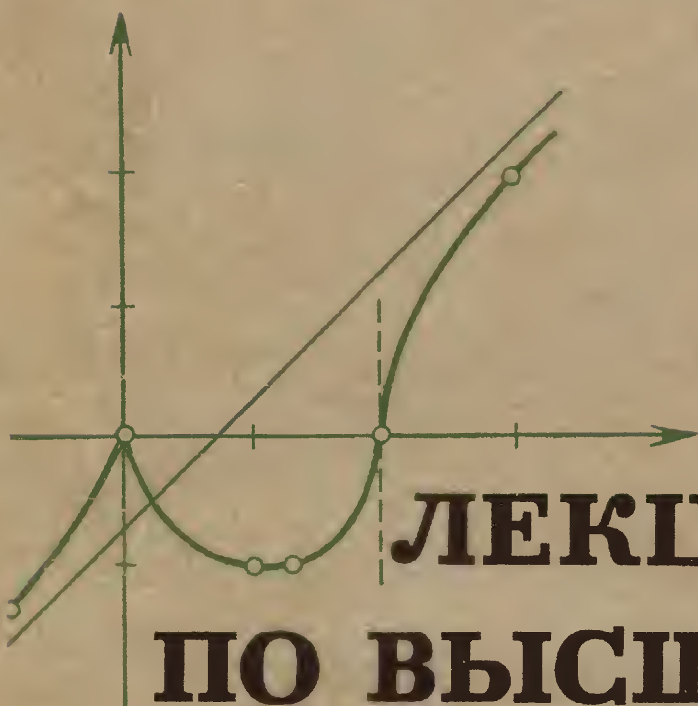
А. Д. МЫШКИС



ЛЕКЦИИ ПО ВЫСШЕЙ МАТЕМАТИКЕ

ЛЕКЦИИ ПО ВЫСШЕЙ МАТЕМАТИКЕ

ИЗДАНИЕ ЧЕТВЕРТОЕ, СТЕРЕОТИПНОЕ

Допущено Министерством
высшего и среднего специального образования СССР
в качестве учебного пособия для студентов
высших технических учебных заведений

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1973

517
М 96
УДК 510 (022)

Анатолий Дмитриевич Мышкис

ЛЕКЦИИ ПО ВЫСШЕЙ МАТЕМАТИКЕ

М., 1973 г., 640 стр. с илл.

Редакторы *Н. В. Воскресенская* и *А. П. Баева*

Техн редактор *А П. Колесникова* Корректор *Н. Д. Дорохова*

Печать с матриц. Подписано к печати 19/XII 1972 г. Бумага 60×90$^{1}/_{16}$, тип № 2 Физ. печ л 40 Условн. печ. л. 40. Уч.-изд. л. 43,74. Тираж 85 000 экз. Цена книги 1 р. 32 к.
Заказ № 659

Издательство «Наука»
Главная редакция физико математической литературы
117071, Москва, В-71, Ленинский проспект, 15

«Союзполиграфпром» при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли Отпечатано в ордена Трудового Красного Знамени Ленинградской типографии № 1 «Печатный Двор» имени А. М Горького, Ленинград, Гатчинская ул., 26 с матриц ордена Трудового Красного Знамени Первой Образцовой типографии имени А. А. Жданова, Москва, М 54, Валовая, 28.

М $\frac{0223-1720}{042(02)-73}$ 18-73

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Эта книга написана на основе лекций, прочитанных автором на протяжении ряда лет студентам высших технических учебных заведений различных специальностей, а также студентам-физикам. Ее содержание соответствует утвержденной в 1964 г. программе общего курса высшей математики для инженерно-технических специальностей вузов. Некоторые менее существенные, по мнению автора, пункты из этой программы в книге опущены. С другой стороны, добавлен ряд вопросов, выходящих из указанной программы, но непосредственно примыкающих к ней. Для удобства читателя изложение этих вопросов напечатано мелким шрифтом; мелким шрифтом набраны также пункты, которые в указанной программе приведены как необязательные, и примеры.

В данном курсе автор стремился отобрать материал и преподнести его так, чтобы наряду с воспитанием необходимого «математического мировоззрения» по возможности облегчить дальнейшее применение математики к специальным дисциплинам. Формальная полнота формулировок и доказательств при этом не являлась самоцелью, так как в приложениях математики эта формальная полнота часто не помогает делу и поэтому в прикладных работах обычно игнорируется. Оговорки сделаны лишь постольку, поскольку они могут на текущем этапе изложения привести к ошибкам в «математическом мировоззрении» или в приложениях; совсем не учитывались факты и объекты, имеющие в настоящее время значение математических патологий. (Например, если упоминаются «все функции», то в их число никогда не включаются неизмеримые по Лебегу функции и даже всюду разрывная функция Дирихле и т. п.) Мы старались, максимально используя интуицию, показать смысл основных математических понятий, убедительно объяснить причину основных математических фактов (считая, что «доказательство» и должно быть таким объяснением) и в возможно большей степени продемонстрировать работающий аппарат. При этом мы сознательно шли на огрубление формулировок и доказательств, применяя доказательство на частных случаях, ссылку на наглядность и т. п. Такой подход, как нам кажется, характерен для современной прикладной математики, основными задачами которой являются наиболее экономные по затраченным усилиям правильное качественное описание фактов и дове-

дение решения поставленной задачи до числа. (Этот подход принципиально отличается от позиций «чистой» математики, которая во главу ставит логическую цельность рассмотрений и разрешает опираться лишь на полностью логически обоснованные положения.) Именно позиции прикладной математики, по мнению автора, должны определять характер преподавания математики инженерам и физикам; впрочем, преподаватель для этого должен хорошо ориентироваться в обеих позициях.

Эту общую установку (изложенную более подробно автором в статье «Что такое прикладная математика?», журнал «Вестник высшей школы», 1967, № 4, стр. 74—80) пока еще трудно осуществить последовательно. Поэтому автор с благодарностью воспримет любые советы и критические замечания читателей, как математиков, так и не математиков.

Автор стремился сделать так, чтобы книгу можно было использовать как при прохождении курса в институте, так и при самообразовании. При помощи разбиения материала на небольшие параграфы и пункты мы старались добиться того, чтобы книгу можно было читать в том или ином объеме, не обязательно подряд, в зависимости от специальности и потребностей. При этом имелось в виду дать возможность пользоваться этой книгой также и заочникам и тем, кто хочет (например, на базе когда-либо прослушанного курса высшей математики) познакомиться с той или иной ее главой. Для этой же цели в ряде мест указана дополнительная литература, список которой приведен в конце книги; ссылки на этот список обозначаются номерами в квадратных скобках. В конце книги приведен также подробный алфавитный указатель; с его помощью легко разыскать определение или разъяснение встретившегося непонятного термина.

При систематическом прохождении курса в институтах материал, относящийся к аналитической геометрии и линейной алгебре, часто выделяется в отдельную дисциплину. Книга допускает такое выделение; при этом указанный материал составляет в точности содержание глав II, VI, VII, X и XI.

Обращаем внимание читателя на способ нумерации, принятый в книге. В каждой главе параграфы, а также пункты и формулы нумеруются подряд, начиная с первого номера. При ссылках номер текущей главы не упоминается: например, в тексте гл. IV выражение «формула (2)» означает «формула (2) гл. IV», а выражение «формула (III. 2)» означает «формула (2) гл. III».

Изучению теоретического материала должны сопутствовать упражнения. Здесь в значительной степени можно ориентироваться на известные сборники упражнений [2], [4], [25], [51]. Правда, некоторые разделы прикладной математики в этих сборниках освещены недостаточно, так что было бы желательно провести работу по подбору интересных и поучительных задач в этих азделах.

Книга может оказаться полезной читателям различных специальностей, сталкивающимся с приложениями математики, не только инженерам, но и физикам и т. д. Конечно, современная прикладная математика содержит множество специальных разделов, значительно выходящих за рамки этого курса. Автор предполагает написать продолжение, содержащее дополнительные главы: теорию функций комплексного переменного, вариационное исчисление, основы математической физики, дополнительные вопросы теории обыкновенных дифференциальных уравнений и другие разделы.

При подготовке второго издания книга была значительно переработана. Это вызвано было прежде всего изменением программы: так, в книгу включен ряд новых вопросов, в том числе новая глава по теории вероятностей, тогда как некоторые разделы выпущены. Кроме того, весь текст заново пересмотрен с учетом критических замечаний многих товарищей, в частности, замечаний, сделанных при обсуждении книги на заседании втузовской секции Московского математического общества. Некоторые места книги написаны или переработаны под влиянием идей и замечаний Л. М. Альтшулера, Я. Б. Зельдовича и Б. О. Солоноуца. Всем этим товарищам автор выражает свою признательность.

19 апреля 1966 г.

А. Д. Мышкис

ВВЕДЕНИЕ

1. Предмет и происхождение математики. Числовые расчеты проникают во все области деятельности инженеров всевозможных специальностей, физиков, химиков и работников многих других профессий. В связи с развитием науки и техники приходится выводить и применять все более сложные законы, решать все более сложные задачи, проводить все более и более сложные подсчеты.

Все эти расчеты основаны на *математике*, науке, которая по определению Ф. Энгельса «имеет своим объектом пространственные формы и количественные отношения действительного мира» ([32], стр. 37), числовые и геометрические соотношения в их всевозможных проявлениях. Математика представляет собой значительный отдел в общей сумме человеческих знаний и приспособлена к обслуживанию самых разнообразных областей науки и практической деятельности.

Все основные математические понятия возникли и развивались в соответствии с потребностями естествознания (физики, механики, астрономии и т. д.) и техники. Положение Энгельса о том, что «уже с самого начала возникновение и развитие наук обусловлено производством» ([32], стр. 500), полностью относится к математике. Само название одного из основных отделов математики — «геометрия» (в переводе — измерение земли) — указывает на происхождение этого отдела из потребностей практики. Энгельс так описывает последовательное развитие отдельных отраслей естествознания: «Сперва *астрономия*, которая уже из-за времен года абсолютно необходима для пастушеских и земледельческих народов. Астрономия может развиваться только при помощи *математики*. Следовательно, приходилось заниматься и математикой. — Далее, на известной ступени развития земледелия и в известных странах (поднимание воды для орошения в Египте), а в особенности вместе с возникновением городов, крупных построек и развитием ремесла развивалась и *механика*. Вскоре она становится необходимой также для *судоходства* и *военного дела*. — Она тоже нуждается в помощи математики и таким образом способствует ее развитию» ([32], стр. 500). И в другом месте: «Как и все другие науки, математика возникла из *практических потребностей* людей: из измерения площадей земельных участков и вместимости сосудов, из счисления времени и из

механики .. *чистая* математика *применяется* впоследствии к миру, хотя она заимствована из этого самого мира и только выражает часть присущих ему форм связей, — и как раз *только поэтому* и может вообще применяться» ([32], стр. 37—38).

Для решения более сложных задач потребовался более сложный *математический аппарат* (т. е математические правила, приемы, формулы и т п), который привел, в частности, к возникновению высшей математики Поэтому не случайно основы высшей математики были созданы в XVII—XVIII вв , т. е. в период начала развития капиталистического производства, хотя элементы ее имелись, например, еще в трудах величайшего математика и механика Древней Греции Архимеда (287—212 гг до н э) В трудах ее основателей — выдающегося французского философа, физика, математика и физиолога Р Декáрта (1596—1650), гениального английского физика, механика, астронома и математика И Нью́тона (1642—1727), немецкого ученого, великого математика и философа-идеалиста Г Лéйбница (1646—1716), великого математика, механика и физика Л Эйлера (1707—1783) и многих других крупнейших ученых — целые отделы математики создаются для анализа явлений природы или для решения технических задач Как и в каждой науке, в математике практика явилась и является основным источником научных открытий Другим основным источником служит «внутриматематическая» потребность систематизации найденных фактов, выяснения их взаимосвязи и т п

2. Значение математики и математических знаний. В настоящее время в естествознании и технике роль математики вообще и высшей математики в частности огромна и на наших глазах стремительно возрастает На математике основаны все области физики, механики, многие отделы других естественных и ряда общественных наук, все инженерные дисциплины, кроме чисто описательных, и т. д Расчет конструкции самолета или плотины гидроэлектростанции, глубокий анализ процессов деформации металла, распространения радиоволн, диффузии нейтронов в атомном реакторе — все это сейчас совершенно невозможно без систематического применения высшей математики

Подлинным триумфом советской науки явился запуск первых в мире искусственных спутников Земли, космических ракет и кораблей, нет сомнения, что среди факторов, которые сделали возможным этот успех, одним из первых явился высокий уровень развития расчетных методов в нашей стране Создание быстродействующих электронных вычислительных машин и других математических автоматов еще более расширяет область непосредственного применения высшей математики и дает возможность проводить важнейшие расчеты там, где они раньше были бы немыслимы Это относится, в частности, к таким областям экономики и организации производства, как разработка *оптимальных* (наилучших) планов размещения капиталовложе-

ний, планов строительства, производства и перевозок, руководство технологическими процессами, диспетчерская служба, где применение современных вычислительных средств уже сейчас приносит ежегодную экономию, исчисляемую в десятках миллионов рублей. В последние годы математика начала эффективно применяться в таких, казалось бы, далеких от нее науках, как биология, языкознание, медицина, география и т. д. Блестяще подтверждается мысль Маркса о том, что наука только тогда достигает совершенства, когда ей удается пользоваться математикой.

Именно поэтому к математическому образованию инженера предъявляются сейчас очень высокие требования он должен твердо знать необходимые положения высшей математики и уметь применять их, когда это требуется. Тогда она превратится для него в мощный инструмент, значительно помогающий ему в его деятельности

Выдающийся русский ученый — математик, механик и кораблестроитель, академик А. Н Крылов (1863—1945) в статье «Значение математики для кораблестроения» писал, что математика для инженера «есть инструмент такой же, как штангель, зубило, ручник, напильник для слесаря... Инженер должен по своей специальности уметь владеть своим инструментом, но он вовсе не должен уметь его делать . слесарь не должен уметь сам насекать напильник, но должен выбрать тот напильник, который ему надо».

А. Н. Крылов сравнивает математика с инструментальщиком, который готовит на склад всевозможные инструменты — от тяжелого молота до микроскопа Инженер, вошедший в этот склад, поражается обилием и разнообразием накопленного за тысячелетия материала Присмотревшись же ближе, « он среди этого бесчисленного разнообразия заметит ряд, видимо, издавна систематически подобранных ассортиментов, остающихся почти неизменными в течение 150 лет, к тому же кладовщик ему подскажет, что их так часто требуют, что и не напасешься, а за остальным заходят лишь знатоки — мастера и любители.

Не отнестись ли ему с доверием к этим, еще издавна великими мастерами подобранным ассортиментам и не следует ли ему воспользоваться этими готовыми и десятилетиями, если не столетиями, испытанными инструментами и научиться ими правильно и искусно владеть, а затем уже, когда он сам станет знатоком и мастером, порыться и в остальных сокровищах и попытаться извлечь из них именно то, что ему надо .

Так вот эти систематические ассортименты — это те курсы, которые вам читают, и те руководства, изучение которых вам рекомендуют, а кладовщики и инструментальщики — это те профессора и руководители, которые вас обучают Может быть, они сами и не инженеры, но зато они хорошо знают и хорошо владеют вверенным инструментом, склад свой они изучили и знают, где и что в нем можно найти».

К этому следует добавить, что за последние годы набор таких «инструментов», применяемых в современной технике, чрезвычайно расширился и многое из того, чем недавно пользовались лишь мастера, получило весьма широкое распространение. Это особенно относится к наиболее передовым областям техники, а также к современным физике, механике и т. п., где требования во владении отдельными разделами математики к специалистам бывают не ниже, чем к математикам.

Надо иметь в виду также, что как огромная часть научно-технической литературы, так и значительная доля специальных предметов насыщены приемами и формулами высшей математики. Если ею владеть слабо, то приходится тратить значительные усилия на то, чтобы разобраться во всех этих формулах, и при этом часто бывает, что основные положения изучаемого вопроса остаются в стороне. Надо владеть математикой настолько, чтобы, говоря несколько образно, перестать ее замечать; а это достигается только на достаточно высокой стадии математического развития. Более того, свободное владение основными математическими понятиями приводит к более глубокому пониманию целого ряда специальных вопросов (колебания, движения сплошных сред и т. д.). «Математика очами физическими по справедливости называться может» (М. В. Ломоносов, «Слово о пользе химии»).

3. Абстрактность. Математика не является инженерной дисциплиной, и ее втузовский курс должен не излагать какие-либо технические вопросы, а дать будущему инженеру или научному работнику необходимое математическое развитие. Поэтому временами материал, рассматриваемый в курсе высшей математики, кажется студенту «слишком абстрактным». Однако абстрактность в математике необходима: она порождается не тем, что математика мало связана с практической деятельностью, а, наоборот, тем, что она приспособлена к самым разнообразным видам этой деятельности. Так, выяснив в геометрии, чему равен объем «абстрактного» цилиндра, мы можем легко найти объем любого конкретного цилиндра, является ли он деталью механизма, колонной или частью пространства, занятой электрическим полем. Аналогично в курсе высшей математики отдельные применения ее законов рассматриваются в качестве примеров; сами же эти законы формулируются в «чистом» виде, без непосредственной связи с каким-либо определенным видом практической деятельности или с какой-либо естественнонаучной или технической дисциплиной (хотя мы отбирали именно те математические вопросы, которые или сами непосредственно применяются, или же необходимы для правильного понимания таковых).

Ф. Энгельс по поводу абстрактности математики писал: «Как понятие числа, так и понятие фигуры заимствованы исключительно из внешнего мира, а не возникли в голове из чистого мышления. Должны были существовать вещи, имеющие определенную форму,

и эти формы должны были подвергаться сравнению, прежде чем можно было прийти к понятию фигуры... Но чтобы быть в состоянии исследовать эти формы и отношения в чистом виде, необходимо совершенно отделить их от их содержания, оставить это последнее в стороне как нечто безразличное; таким путем мы получаем точки, лишенные измерений, линии, лишенные толщины и ширины, разные a и b, x и y, постоянные и переменные величины» ([32], стр. 37).

Таким образом, математика изучает «чистые», «идеальные» (схематизированные) формы, связи, процессы и т. п., которые лишь приближенно реализуются в действительности: например, реальный цилиндр никогда не является точным математическим цилиндром. Здесь особенно ярко проявляется черта, свойственная всякому познанию: из бесконечного многообразия свойств, характеризующих тот или иной реальный объект, процесс и т. п., выбирать и рассматривать лишь небольшое число свойств, отвлекаясь (*абстрагируясь*) от остальных, как от несущественных в данном рассмотрении. Однако в некоторых случаях бывает так, что предположения (допущения, гипотезы) о несущественности этих остальных свойств не оправдываются, и тогда вывод, полученный математически, может прийти в противоречие с действительностью; об этой роли допущений всегда надо помнить при применении математики.

«Чистота» рассматриваемых форм и связей в математике, как ни в какой другой науке, влечет за собой особенно большое значение логических выводов, которое хорошо известно уже из курса элементарной математики. Так и в высшей математике утверждения должны доводиться до полной ясности и надлежащим образом обосновываться, чтобы на них можно было без опасения опираться как на объективные (не зависящие от нас) истины, правильно отображающие действительность. Конечно, выводы, сделанные на основе экспериментов, наблюдений и аналогий, встречаются и в математике, причем в высшей математике не так уж редко, однако все же существенно реже, чем в других науках.

В математике даже существует известная тенденция выводить все положения чисто логически из немногих отправных утверждений — аксиом; это — так называемый *дедуктивный метод*. В нашем курсе, предназначенном для прикладников, мы не всегда будем пунктуально придерживаться этого метода, и тем, кто сочтет то или иное рассуждение не вполне совершенным с точки зрения логики, заинтересуется исключениями из общих правил и вообще захочет более глубоко овладеть материалом, следует обратиться к курсам, написанным для математиков, например, [48]. Чтобы посмотреть на один и тот же вопрос с разных сторон, полезно привлечь также другие втузовские курсы ([5], [19], [38], [43]), из которых мы особенно рекомендуем книгу А. Ф. Берманта.

4. Характерные черты высшей математики. Резкую грань между математикой элементарной, изучаемой в основном в средней

школе, и математикой высшей, изучаемой в высшей школе, провести невозможно. Сейчас разделение математики на элементарную и высшую весьма условно и определяется не научными причинами, не тем, что это действительно различные области математики, а в значительной степени историческими причинами, тем, что «элементарная математика» и «высшая математика» (в обиходном понимании) возникли в разные эпохи, имеют разный объем и изучаются в разных типах школ. Тем не менее можно отметить несколько черт, характерных для высшей математики (взятой, например, в объеме нашего курса).

Одной из отличительных черт высшей математики является универсальность, общность ее методов. Рассмотрим, например, задачу о вычислении объемов тел. Элементарная математика дает формулы для вычисления объемов призмы, пирамиды, конуса, цилиндра, шара и некоторых других простых тел. Для вывода каждой из этих формул требовалось отдельное рассуждение, иногда довольно сложное. В высшей же математике даются единые формулы для объема любого тела, для длины любой линии, площади любой поверхности и т. п. В качестве другого примера рассмотрим задачу о движении точки, на которую действуют заданные силы. В школьном курсе физики рассматриваются только равномерное прямолинейное, равноускоренное (или равнозамедленное) прямолинейное и равномерное круговое движения, и при помощи средств элементарной математики исследовать другие виды движений затруднительно. Методы же высшей математики дают возможность исследовать любые виды движений, которые могут встретиться на практике.

В связи с указанной чертой находится и другая характерная черта высшей математики — систематическое рассмотрение переменных величин. При исследовании процессов, объектов средствами элементарной математики обычно важнейшие факторы — скорости или ускорения, плотности, массы, силы и т. п. — считаются постоянными, да и то при этом исследование удается провести лишь в особенно простых случаях. Если же указанные величины меняются существенно (как это часто бывает) и их никак нельзя принять за постоянные, то чаще всего и применяется высшая математика.

Ф. Энгельс писал: «Лишь дифференциальное исчисление дает естествознанию возможность изображать математически не только *состояния*, но и *процессы*: движение» ([32], стр. 587). А дифференциальное исчисление является одним из первых разделов высшей математики, в котором решаются разнообразные задачи, связанные в основном с изучением зависимости одних величин от изменения других. Так как величины и зависимости между ними могут быть самые разнообразные (ускорения, скорости и пути; плотности, массы и силы и т. п.), то дифференциальное исчисление широко проникает во всевозможные естественные и технические науки.

Третьей характерной чертой высшей математики являются тесная взаимосвязь ее различных разделов и систематическое объединение

вычислительных, аналитических (формульных) и геометрических методов (в отличие от элементарной математики, где связь алгебры с геометрией была довольно эпизодической). На основе метода координат геометрические задачи приводятся к решению алгебраических уравнений, графики применяются для иллюстрации зависимости между переменными величинами, аналитические методы интегрального исчисления — для вычисления площадей и объемов и т. д.

5. Замечания о развитии математики. Некоторые исторические сведения мы будем сообщать в течение курса по мере изложения материала. Однако и здесь следует сделать несколько вводных замечаний, которые полезно иметь в виду. Значительная часть «общих» отделов высшей математики, составляющих основу обязательной программы для инженеров многих специальностей, — метод координат, дифференциальное и интегральное исчисления и т. д. — была создана в XVII — XVIII вв. Эта весьма важная часть применяется и потому излагается сейчас примерно в таком виде, какой она приняла после работ Л. Эйлера. В XIX и XX вв. математика продолжала бурно развиваться, было создано большое число «специальных» отделов, изучаемых сейчас в зависимости от специальности; кроме того, в этот период старые отделы получили более полное логическое обоснование и приняли тот вид, в котором они сейчас излагаются для математиков.

Математика создавалась и создается усилиями ученых многих стран. Значительный труд в это дело вложили и наши отечественные ученые. В «специальных» отделах, особенно самых современных, их роль огромна и очевидна как в теории, так и в ее приложениях. В более же старых отделах, созданных в своей основе до прочного и постоянного обоснования русской математической школы на мировой арене (которое начинается примерно с двадцатых годов XIX в.), имена этих ученых, естественно, встречаются реже. Однако непременно надо отметить Л. Эйлера, швейцарца по происхождению, бóльшая часть научной деятельности которого протекала в стенах Петербургской Академии наук и бóльшая часть научных трудов которого (473 из 865!) была напечатана впервые в изданиях этой Академии. (Эйлер и умер в Петербурге; прах его находится в ленинградском некрополе.) Его выдающиеся результаты в самых разнообразных отделах математики, механики, физики и других наук лежат в самой основе этих отделов.

После Л. Эйлера крупнейшим в России был великий математик, создатель неевклидовой геометрии, мыслитель-материалист Н. И. Лобачевский (1792 — 1856). Он впервые детально показал, что евклидова геометрия (которую мы изучаем в школах) не является единственно возможной, и построил новую, неевклидову геометрию, которую теперь называют геометрией Лобачевского. Это открытие имело огромное философское значение, а кроме того, знаменовало переход всей математики к ее новому виду, характерному для современного

периода, когда большое внимание уделяется точным логическим выводам, анализу аксиом и т. д. Отметим, что, согласно общей теории относительности, физическое пространство является евклидовым лишь приближенно, хотя его «неевклидовость» и не совсем такова, как представлял себе Н. И. Лобачевский. Лобачевскому принадлежат также важные работы и в других областях математики; кроме того, он был выдающимся общественным деятелем. От Лобачевского ведет начало развитие математики в Казани.

Интенсивное и постоянное развитие математики в Петербурге началось работами выдающегося математика, одного из основателей петербургской математической школы, академика М. В. Остроградского (1801—1862). Однако подлинным создателем этой школы — коллектива ученых, имевшего и имеющего мировое значение, был великий русский математик и механик, академик П. Л. Чебышёв (1821—1894). Он был автором весьма важных результатов в различных отделах математики и ее приложений к теории механизмов, к картографии и т. д. Для его творчества характерно очень верное высказывание, которое он сделал в статье «Черчение географических карт»: «Сближение теории с практикой дает самые благотворные результаты и не одна только практика от этого выигрывает; сами науки развиваются под влиянием ее; она открывает им новые предметы для исследования или новые стороны в предметах, давно исследованных... наука находит себе верного руководителя в практике».

Виднейшими представителями петербургской математической школы были, в частности, ученики П. Л. Чебышёва — выдающийся математик, академик А. А. Марков (1856—1922) и выдающийся математик и механик, академик А. М. Ляпунов (1857—1918). Бессмертной заслугой первого является создание теории случайных процессов, а второго — теории устойчивости движения; оба эти отдела математики интенсивно развиваются и сейчас и имеют большое прикладное значение.

Во второй половине XIX в. и позже математика начинает развиваться в Киеве, Москве, Одессе, Харькове и других центрах страны.

6. Математика в Советском Союзе. Уже в дореволюционной России трудами отдельных ученых математика была поставлена на высокий уровень, был достигнут ряд выдающихся результатов, заложены здоровые традиции. Однако подлинный расцвет математики наступил у нас после Великой Октябрьской социалистической революции, что, конечно, объясняется коренными сдвигами в экономике и культуре нашего народа. Вместе с развитием прогрессивных черт науки дореволюционного периода математика в Советском Союзе приобрела ряд новых характерных особенностей.

Прежде всего несравненно увеличилось число творчески работающих математиков, объединенных в многочисленные коллективы на базе научно-исследовательских и учебных институтов. У нас сейчас имеются десятки выдающихся математиков; это — академики

А. Д. Александров, П. С. Александров, С. Н. Бернштейн, Н. Н. Боголюбов, И. Н. Векуа, И. М. Виноградов, В. М. Глушков, Л. В. Канторович, М. В. Келдыш, А. Н. Колмогоров, М. А. Лаврентьев, Ю. В. Линник, А. И. Мальцев, Н. И. Мусхелишвили, П. С. Новиков, И. Г. Петровский, Л. С. Понтрягин, В. И. Смирнов, С. Л. Соболев и А. Н. Тихонов, а также многие другие ученые, трудами которых наша страна по развитию математики выведена на одно из первых мест в мире. Ими получены результаты первостепенного значения во всех актуальных областях математики (а не только в избранных областях, как это было ранее). Математические коллективы создаются и укрепляются по всей территории Советского Союза, как в старых математических центрах, указанных ранее, так и в новых — в Баку, Вильнюсе, Воронеже, Горьком, Ереване, Львове, Минске, Новосибирске, Ростове, Саратове, Свердловске, Ташкенте, Тбилиси и других городах. Значительно возросло издание математической литературы — научной и учебной, отечественной и переводной.

Резко усилилась связь математики с другими науками (особенно, с физикой, механикой и т. д.), с техникой и народным хозяйством, что оказывается плодотворным как для математики, так и для областей, в которых она применяется. Многие математики решают прикладные задачи и развивают теоретические вопросы, необходимые для этого решения; этому немало должно также способствовать создание многочисленных вычислительных центров, оборудованных современной техникой. С другой стороны, многие физики, механики, специалисты в области технических наук и другие ученые развивают важные для них отделы математики и виртуозно применяют их в своей специальности. Так, образец органического объединения математики и ее приложений и вообще теории и практики показали великий русский ученый, основоположник современной аэро- и гидромеханики Н. Е. Жуковский (1847—1921), которого В. И. Ленин назвал «отцом русской авиации», А. Н. Крылов, выдающийся советский ученый в области теоретической механики, аэро- и гидромеханики, академик С. А. Чаплыгин (1869—1942), а также ряд других ученых.

Весьма важной стороной является также распространение математики вширь, что связано с улучшением качества среднего и высшего образования, со значительным расширением контингента обучающихся и преподавателей, с глубоким проникновением математических методов в инженерную деятельность, в физику, механику и т. д. Нет сомнения, что этот процесс овладения математикой многими людьми продолжится еще далее, а роль математики во всей нашей жизни еще более усилится.

ВЕЛИЧИНА И ФУНКЦИЯ

§ 1. Величина

1. Понятие величины. Понятие величины настолько широко и всеобъемлюще, что ему трудно дать точное определение. Массы, давления, работы, заряды, длины и объемы, целые и дробные числа — все это примеры величин. На первой стадии *величиной можно считать то, что, выраженное в определенных единицах* (например, масса — в граммах или тоннах и т. п.), *характеризуется своим численным значением.* Так, площадь круга является величиной, поскольку она, выраженная, например, в квадратных сантиметрах, полностью характеризуется своим численным значением (5, π и т. п.); сам круг, конечно, не является величиной, так как для него характерна определенная *форма*, которая не выражается каким-либо числом.

За последние годы многие понятия, ранее воспринимавшиеся лишь качественно (такие, например, как эффективность, количество информации и даже степень правдоподобия), «повышены в должности» и переведены в разряд величин. Каждый такой перевод является радостным событием, так как он дает возможность применить к указанным понятиям количественный математический анализ, что часто оказывается очень эффективным.

2. Размерность величины. *Размерностью величины называется та единица, через которую эта величина выражена.* Так, размерностью массы обычно служит грамм или килограмм; размерностью площади — квадратный сантиметр или квадратный метр и т. п. Размерность обозначается квадратными скобками; например, если M — масса, S — площадь, то в международной системе единиц $[M]=кг$ (т. е. килограмм), $[S]=м^2$ (т. е. квадратный метр).

Обычно размерности некоторых величин принимаются за основные, а размерности остальных величин выражаются через эти основные. Так, в механике в международной системе единиц размерности длины (*м*), массы (*кг*) и времени (*сек*) считаются основными; через них выражаются, например, размерности скорости (*м/сек*) или силы ($кг \cdot м/сек^2$).

Складывать или вычитать можно только величины одинаковой размерности, причем размерность суммы такая же, как размерность слагаемых. Множить или делить друг на друга можно величины любой размерности; при умножении (*или делении*) *величин их размерности тоже множатся или соответственно делятся*).

Часто рассматриваются величины *безразмерные* («отвлеченные»). Так, *отношение двух величин одинаковой размерности является безразмерным.* Численное значение величины, которое является отношением этой величины к ее выбранной единице, также безразмерно; например, численным значением массы в 5 *кг* служит «безразмерная масса» 5. Безразмерную массу можно получить также, взяв отношение изучаемой массы к некоторой *характéрной* в рассматриваемом процессе массе (хорошо известной и принимаемой в данном процессе за эталон для сравнения). Подобным образом вводятся безразмерные длина, время и т. п.

В курсе математики величины обычно считаются безразмерными. Безразмерная величина полностью характеризуется своим численным значением, ее «единицей» служит число 1.

3. Постоянные и переменные величины. *Величина, участвующая в некотором рассмотрении, может либо принимать различные значения, либо принимать одно определенное значение; в первом случае она называется переменной величиной, а во втором — постоянной (константой).* Так, при рассмотрении воды в бассейне давление в различных точках есть величина переменная, оно зависит от места замера, тогда как плотность в разных точках можно с достаточной точностью считать величиной постоянной. Другой пример: при рассмотрении процесса сжатия определенной порции газа при постоянной температуре давление и объем будут величинами переменными, а масса и температура — постоянными. Впрочем, надо иметь в виду, что в любом реальном процессе и эти две последние величины несколько меняются, и только если это изменение незначительно и несущественно для остального, можно условно, схематизируя процесс, принять их за постоянные. И в других случаях постоянство тех или иных величин обычно является лишь условным; об этом надо время от времени вспоминать, так как если считать постоянной величину, изменение которой невелико, но существенно для рассмотрения, то можно прийти к ошибочным выводам (что неоднократно бывало).

Величина, постоянная в одном рассмотрении, может в другом аналогичном (похожем) рассмотрении принимать другое значение или даже быть переменной. Такие постоянные величины называются парáметрами данного рассмотрения; они являются его характеристиками. Так, в процессе изотермического сжатия газа масса и температура служат параметрами. При выборе электрической лампочки ее параметрами служат сопротивление, напряжение в сети, на которое она рассчитана, и потребляемая мощность. Конечно, здесь имеются и другие параметры, которые иногда приходится принимать во внимание (например, габариты), но обычно именно эти считаются основными; и в других случаях важно уметь выделить из всевозможных параметров, характеризующих тот или иной объект или процесс, основные, наиболее важные параметры.

4. Числовая ось. Логарифмическая линейка. Величины можно наглядно изображать при помощи *числовой оси*. Чаще всего применяется *прямолинейная ось с равномерной шкалой*. Для ее построения надо выбрать прямую линию, а на ней — начало отсчета, которое обычно обозначается буквой O. Затем надо на этой прямой выбрать одно из двух направлений за положительное (это направление обозначается стрелкой; рис. 1) и принять некоторый отрезок за единицу масштаба. Откладывая этот отрезок от начала отсчета в обоих направлениях, получим точки, отвечающие целым численным значениям рассматриваемой величины. Между этими «целыми точками» расположены точки, отвечающие нецелым значениям — дробным рациональным $\left(\text{например, } \frac{1}{2},\ -2{,}03 \text{ и т. п.}\right)$ и иррациональным $\left(\text{т. е. нерациональным, например } \frac{\sqrt{2}+1}{3},\ -\pi \text{ и т. п.}\right)$. Если рассматриваемая величина размерная (именованная), то и отрезок, принятый за единицу, получает соответствующее наименование: например, на рис. 1 на оси изображаются численные значения времени t, выраженного в секундах; там же изображены точки N ($t = -1$ *сек*), O ($t = 0$ *сек*), M ($t = 1{,}37$ *сек*).

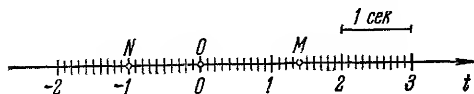


Рис. 1.

Каждому значению величины отвечает некоторая точка на оси, и, наоборот, каждая точка на оси соответствует некоторому значению этой величины. (Здесь и далее имеются в виду, конечно, лишь величины, принимающие вещественные, т. е. действительные значения; о том, как изображаются комплексные величины, мы поговорим в п. VIII. 1.) Поэтому часто значения величины и изображающие точки как бы отождествляются, и говорят просто: дана точка $t = 1{,}37$ *сек* и т. п.

Рис. 2.

Если величина переменная, то она изображается точкой, которая может занимать на оси различные положения (например, которая перемещается с течением времени); такая точка называется *текущей*. Постоянной величине отвечает точка, не меняющая своего положения.

На практике начало отсчета и единицу масштаба выбирают так, чтобы наилучшим образом изобразить диапазон (промежуток) изменения рассматриваемой величины; при этом часто бывает, что само начало на рисунок не попадает, так как изображается, конечно, лишь часть оси. Например, на рис. 2 изображена ось значений длины стержня при его нагревании.

Иногда бывает удобно применять неравномерные шкалы. Так, часто применяется *логарифмическая шкала* (рис. 3), в которой число $n>1$ изображается точкой, полученной откладыванием от некоторой выбранной точки A в положительном направлении отрезка длины $k\lg n$, где k — некоторый выбранный коэффициент пропорциональности. Положительные числа $n<1$ получаются на логарифмической шкале откладыванием от A отрезка $k|\lg n|$ в отрицательном направлении, так как для таких n будет $\lg n<0$.

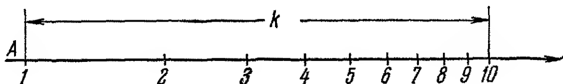


Рис. 3.

Логарифмическая шкала применена, в частности, в *логарифмической линейке*. Чтобы понять принцип ее действия, допустим, что две одинаковые логарифмические шкалы, одна из которых нанесена на корпус линейки, а другая на ее движок, сдвинуты друг относительно друга (рис. 4). Пусть при этом какие-то отметки a и b на

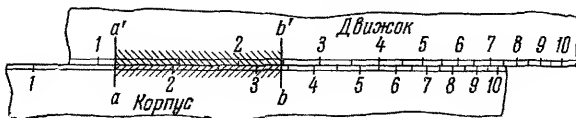


Рис. 4.

нижней шкале совпали с соответствующими отметками a' и b' на верхней шкале. Тогда, исходя из равенства длин заштрихованных отрезков, получим

$$k\lg b-k\lg a=k\lg b'-k\lg a',$$

откуда после простых преобразований (проверьте!)

$$\frac{b}{a}=\frac{b'}{a'}. \qquad (1)$$

Задавая три из величин a, b, a', b', мы сможем на линейке прочитать значение четвертой из них, удовлетворяющей соотношению (1). Например, если взять $a'=1$, мы получим $b=ab'$ или $a=\frac{b}{b'}$. Значит, чтобы найти произведение заданных чисел a и b', надо отметку 1 на движке совместить с отметкой a на корпусе, после чего против отметки b' на движке прочитать на корпусе значение произведения. (А как найти частное заданных чисел?) Иногда удобнее полагать не $a'=1$, а $b'=10$, сдвигая движок не направо относительно корпуса, а налево. Логарифмическая линейка, созданная в XVII в., сейчас вошла в широкий обиход и значительно

облегчает жизнь огромному числу техников, инженеров, физиков и т. д. С помощью дополнительных шкал на ней можно производить ряд дополнительных действий, в том числе извлечение корня, возведение в степень, тригонометрические вычисления, логарифмирование и потенцирование, решение различных уравнений и т. д. Имеется ряд руководств по работе на логарифмической линейке, из которых наиболее известны книжки [37], [42].

В некоторых случаях применяются также криволинейные шкалы (см., например, п. IX. 1). Однако мы в нашем курсе будем всегда, если не оговорено иное, для изображения величин пользоваться прямолинейными осями и равномерными шкалами.

5. Характеристики переменных величин. *Переменная величина, которая принимает сплошь все числовые значения или все значения, заключенные между некоторыми границами, называется непрерывной.* В противоположность этому *величина, принимающая отдельные, «оторванные» друг от друга значения, называется дискрéтной.*

Совокупность тех значений, которые может принимать данная переменная величина, принято называть областью изменения этой величины. Для указания этой области полезно ввести понятие интервала.

Конечным (ограниченным) интервалом называется совокупность всех чисел, заключенных между какими-либо двумя данными числами a и b, которые называются концами этого интервала. При этом сами концы a и b или причисляются к интервалу, или нет, о чем иногда необходимо указывать. Тогда говорят соответственно о *замкнутом* интервале (если концы к нему причисляются) или об *открытом* (в противном случае). На числовой оси конечные интервалы изображаются отрезками.

Интервалы также могут быть *неограниченными* (*бесконечными*) с одной стороны или с обеих сторон. Например, если величина x может принимать любые значения, бóльшие некоторого постоянного числа a, то область изменения величины можно записать неравенствами $a < x < \infty$. Эта область представляет собой неограниченный интервал; этот интервал, конечно, не имеет правого конца, однако условно говорят, что в данном случае правый конец находится на бесконечности. Про такой интервал говорят, что он неограничен *сверху*, так как возрастание величины ассоциируется (связывается в обычном представлении) с ее «повышением»; если ось x проходит слева направо, то интервал $a < x < \infty$ можно также назвать неограниченным справа. Интервалом, неограниченным с обеих сторон, является совокупность всех чисел (геометрически — вся числовая ось).

Областью изменения непрерывной величины служит интервал или совокупность некоторого числа интервалов. Например, если треугольник ABC деформируется всевозможными способами, то угол A будет непрерывной переменной величиной, областью изменения которой будет интервал $0 < \angle A < \pi$, если брать численное значение

угла, выраженного в радианной мере; областью изменения площади S будет интервал $0 < S < \infty$ (конечно, и здесь имеется в виду численное значение площади в определенных единицах, о чем мы впредь не будем все время упоминать). *Областью изменения дискретной величины служит совокупность конечного или бесконечного количества отдельных чисел* или, как говорят геометрически, отдельных точек (но не целых интервалов). Например, какой-либо номер может принимать значения 1, 2, ...; он будет дискретной переменной величиной.

Если переменная величина в некотором процессе меняется все время в одном направлении, т. е. все время возрастает или все время убывает, *она называется монотонной*. Точка, изображающая монотонную величину на числовой оси, движется все время в одну и ту же сторону.

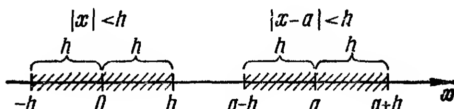


Рис 5.

Чтобы не рассматривать постоянные величины отдельно, *можно считать постоянную величину частным случаем переменной*, которая в силу каких-то причин все время принимает одно и то же значение (ведь и в механике состояние покоя считается частным случаем движения). В этом случае область изменения состоит всего лишь из одной точки.

Величина называется *ограниченной сверху* (или справа), если в процессе своего изменения она все время остается меньше некоторой постоянной величины. Аналогично определяется *ограниченность снизу* (или слева). Величина, ограниченная сверху и снизу, называется просто *ограниченной*.

При исследовании величин часто применяется понятие *абсолютного значения* величины. Как известно из средней школы, это понятие определяется так:

$$|a| = a, \text{ если } a \geqslant 0; \quad |a| = -a, \text{ если } a < 0.$$

Например, $|5| = 5$; $|0| = 0$; $|-5| = 5$ [т. е. $|-5| = -(-5)$].

Абсолютные значения имеют следующие простые свойства: $|a+b| \leqslant |a| + |b|$, причем это неравенство является строгим, если a и b имеют противоположные знаки, в противном же случае обращается в равенство;

$$|ab| = |a| \cdot |b|; \quad \sqrt{a^2} = |a|.$$

Забвение последней формулы, которой в средней школе уделяется недостаточное внимание, служит постоянным источником ошибок.

Значение $|a-b| = |b-a|$ равно расстоянию между точками a и b на числовой оси. Неравенство $|x| < h$ определяет интервал $-h < x < h$, а неравенство $|x-a| < h$ — интервал $-h < x-a < h$, т. е. $a-h < x < a+h$ (на рис. 5 эти интервалы заштрихованы).

§ 2. Приближенные значения величины

6. Понятие приближенного значения. Обычно говорить об абсолютно точном численном значении физической величины невозможно. Например, мы никогда не можем знать длину какого-либо реального тела абсолютно точно. Это происходит не только из-за несовершенства измерения, но также и из-за несовершенства формы самого тела, в результате чего невозможно указать точно, от какой точки и до какой надо производить отсчет (а если вспомнить, что тело состоит из молекул, которые все время движутся, то положение еще осложнится). Более того, в громадном большинстве случаев указание длины с чрезмерно большой степенью точности нецелесообразно, даже если оно возможно при современном уровне измерительной техники. Например, при обмере или проектировании жилого дома было бы нелепо указывать размеры с точностью до сотых долей миллиметра. То же можно сказать о массах, давлениях и т. п. Поэтому численные значения почти всех величин в физике и технике (например, всех непрерывных величин) задаются приближенно.

Математические действия над приближенными значениями величин называются приближенными вычислениями. К настоящему времени создана целая наука о приближенных вычислениях, с рядом положений которой мы познакомимся в дальнейшем. В нашей стране инициатором развития приближенных вычислений был А. Н. Крылов. Его книга «Лекции о приближенных вычислениях», изданная впервые в 1911 г., сохранила значение и в настоящее время.

Выбор степени точности, с которой производится изготовление какой-либо детали, или измерение, или вычисление, — это чрезвычайно ответственное дело. При этом выборе приходится руководствоваться многими соображениями — потребностями, техническими возможностями, экономичностью и т. п.

7. Погрешности. Пусть точное значение какой-либо величины равно A, а приближенное равно a. *Тогда погрешность*, т. е. *отклонение точного значения от приближенного*, равна $A-a$; она может получиться как положительной, так и отрицательной. Эта погрешность обычно бывает точно неизвестна, так как неизвестно значение A. Поэтому обычно задаются *предельные погрешности* α_1 и α_2, между которыми содержится истинная погрешность:

$$\alpha_1 < A - a < \alpha_2, \text{ т. е. } a+\alpha_1 < A < a+\alpha_2.$$

В этом случае говорят, что задана *двусторонняя оценка* величины A. Таким образом, например, формула для длины $L = 9^{+0,2}_{-0,1}$ *мм* означает, что истинное значение длины заключено между $9-0,1=8,9$ *мм* и $9+0,2=9,2$ *мм*.

Так как задавать две предельные погрешности не всегда удобно, то часто задается *предельная абсолютная погрешность* α, т. е.

величина, бо́льшая абсолютного значения погрешности:

$$|A-a|<\alpha, \text{ т. е. } -\alpha<A-a<\alpha \text{ или } a-\alpha<A<a+\alpha.$$

Пусть, например, при измерении некоторой длины l получилось 137 *см*, причем мы можем ручаться за точность до 0,5 *см*. Это значит, что в данном случае $\alpha=0{,}5$ *см* и $136{,}5<l<137{,}5$ *см*; можно написать $l=(137\pm 0{,}5)$ *см*.

Предельная абсолютная погрешность не полностью характеризует точность измерения: например, если она равна 1 *см*, то еще неясно, грубая это ошибка или нет, так как важно, что измеряли — длину кита или длину жука. Качество измерения больше характеризуется *предельной относительной погрешностью* δ, которая вычисляется по формуле

$$\delta=\frac{\alpha}{a}.$$

Предельная относительная погрешность безразмерна и часто выражается в процентах, причем для упрощения ее значение обычно округляется в сторону увеличения. Скажем, в приведенном примере с вычислением длины l предельная относительная погрешность в процентах равна $\frac{0{,}5\cdot 100}{137}=0{,}36\ldots<0{,}4$, т. е. можно сказать, что измерение произведено с предельной относительной погрешностью в $0{,}4\%$ $\left(\text{или даже просто в } \frac{1}{2}\%\right)$.

Для многих прикидочных расчетов достаточна точность (т. е. предельная относительная погрешность) порядка процентов и даже десятков процентов. С другой стороны, например, точное измерение частоты электромагнитных колебаний, являющееся основой автоматического управления космическими кораблями, производится кварцевыми и атомными часами с погрешностью, соответствующей ошибке хода этих часов в одну стотысячную долю секунды в сутки (подсчитайте предельную относительную погрешность!).

8. Запись приближенных чисел. Запись *приближенных чисел*, т. е. приближенных численных значений величин, производится так, чтобы сам вид записи говорил о степени их точности. Обычно их записывают так, что *все цифры верны, кроме последней, сомнительной, в которой допускается ошибка не больше чем на единицу*; впрочем, если ошибка чуть больше, то особенно не придираются. Например, выражение для сопротивления $R=1{,}35\,\Omega$ означает, что $\alpha_R=0{,}01\,\Omega$, т. е. на самом деле $1{,}34<R<1{,}36\,\Omega$. Между формулами $R=1{,}35\,\Omega$ и $R=1{,}3500\,\Omega$ огромная разница, так как эти записи говорят, что первое вычисление производилось с точностью до 0,01, а второе — до $0{,}0001\,\Omega$. (Иногда говорят, что во втором случае точность на два *порядка* выше, или что погрешность на два порядка меньше, чем в первом.) Если при вычислении получилось значение $R=2{,}37\,7\,\Omega$, но уже третья цифра сомнительна или

четвертая нас не интересует, то надо произвести *округление*, т. е. написать $R = 2{,}38\ \Omega$.

Число знаков после запятой говорит о предельной абсолютной погрешности; о предельной же относительной погрешности говорит общее *число верных знаков*, к которым не относят передние нули: например, числа 2,57, 1,7100, 0,015, 0,00210 имеют соответственно 3, 5, 2, 3 верных знаков. Чем больше верных знаков в числе, тем меньше предельная относительная погрешность.

Следует избегать записей вида $M = 1800$ *г*, так как они зачастую не показывают точности измерения (или вычисления). Если вторая цифра сомнительна, следует писать $M = 1{,}8 \cdot 10^3$ *г*, а если четвертая — то $1{,}800 \cdot 10^3$ *г*. Строго говоря, запись $M = 1800$ *г* должна означать, что предельная абсолютная погрешность равна 1 *г*. Это правило не всегда соблюдается, поэтому могут возникнуть недоразумения.

9. Сложение и вычитание приближенных чисел. Рассмотрим пример. Пусть бутылка и пробка взвешивались раздельно, причем массы их оказались соответственно равными $M = 323{,}1$ *г* и $m = 5{,}722$ *г* (пробка взвешивалась на более точных весах). Для нахождения суммарного веса бутылки с пробкой было бы неправильно считать так:

$$\begin{array}{r} M = 323{,}1 \\ m = 5{,}722 \\ \hline M + m = 328{,}822\ \text{г}. \end{array}$$

Действительно, вес бутылки определен только с точностью до 0,1 *г*, и потому сотые и тысячные в ответе являются не только лишними цифрами, но даже вредными: форма ответа такова, как будто $M + m$ определено с точностью до 0,001, что неверно. Поэтому при сложении m следует округлить до 0,1, т. е. проводить вычисления так:

$$\begin{array}{r} M = 323{,}1 \\ m = 5{,}7 \\ \hline M + m = 328{,}8\ \text{г}; \end{array}$$

этот же ответ получится, если округлить результат, подсчитанный выше. Таким образом, в сумме берется столько знаков после запятой, сколько их имеется у слагаемого с наибольшей абсолютной погрешностью.

Если слагаемых много, то ошибки в них могут складываться и дать большую ошибку в сумме (систематический «недолив»). В таких случаях рекомендуется *правило лишнего знака*: оставлять один лишний знак, а в ответе произвести его округление.

Пусть, например, надо найти сумму

$$K = 132{,}7 + 1{,}274 + 0{,}06321 + 20{,}96 + 46{,}1521.$$

Самая большая абсолютная погрешность у первого слагаемого: она равна 0,1. Поэтому прочие слагаемые округляем до 0,01:

$$132{,}7+1{,}27+0{,}06+20{,}96+46{,}15=201{,}14,$$

т. е. $K=201{,}1$. Если бы мы не воспользовались правилом лишнего знака и округляли все слагаемые до 0,1, то получили бы менее точный результат:

$$K=132{,}7+1{,}3+0{,}1+21{,}0+46{,}2=201{,}3.$$

Другой пример. Пусть надо найти сумму $N=\sqrt{5}+\sqrt{6}+\sqrt{7}+\sqrt{8}$ с точностью до 0,01, причем считается, что целые числа, стоящие под знаками радикала, совершенно точные. Пользуясь правилом лишнего знака, берем из таблиц значения корней с точностью до 0,001:

$$2{,}236+2{,}449+2{,}646+2{,}828=10{,}159,$$

т. е. $N=10{,}16$.

Если число слагаемых весьма велико, скажем несколько сотен, следует пользоваться двумя лишними знаками.

При вычислении суммы нескольких слагаемых, заданных с одинаковым числом знаков после запятой, следует иметь в виду, что предельная абсолютная погрешность у суммы будет больше, чем у слагаемых; поэтому ответ целесообразно округлить до предыдущего знака. Например, пусть

$$L=1{,}38+8{,}71+4{,}48+11{,}96+7{,}33.$$

Складывая, получим $L=33{,}86$. Однако последняя цифра очень сомнительная; поэтому следует написать ответ в виде $L=33{,}9$.

Предельная абсолютная погрешность суммы или разности нескольких величин равна сумме предельных абсолютных погрешностей этих величин. Например, если две величины определены с точностью до 0,1, то, как легко понять, сумма или разность этих величин определены с точностью до 0,2, так как ошибки могут сложиться. Если слагаемых много, то очень маловероятно, чтобы все ошибки сложились. В этом случае для определения погрешности суммы надо пользоваться методами теории вероятностей (см. п. XVIII.15). Из них вытекает, что один знак в сумме надо округлять, как это было сделано при вычислении L, начиная примерно с пяти слагаемых, а два знака — примерно с 500.

При вычитании приближенных чисел правила те же, что при сложении, но надо дополнительно иметь в виду, что при вычитании близких чисел относительная точность резко ухудшается. Например, пусть надо найти $P=327{,}48-326{,}91$. В вычитаемом и уменьшаемом $\alpha=0{,}01$, т. е. $\delta<\frac{0{,}01}{300}\cdot 100\%<0{,}004\%$.

В разности же $P=0{,}57$ предельная абсолютная погрешность равна 0,02, поэтому предельная относительная погрешность $\delta_P=\frac{0{,}02}{0{,}57}\cdot 100\%=3{,}5\%$.

Относительная погрешность увеличилась в 1 000 раз!

Поэтому надо стараться измерять или вычислять разности близких чисел непосредственно, без выполнения такого вычитания: не следует вычислять вес шляпы, взвесившись сначала в шляпе, а затем без нее. Формулы же, содержащие разности близких величин, надо стараться преобразовать, избавляясь от таких разностей, если они могут существенно нарушить точность вычислений. Например, выражение вида $Q=a-\sqrt{a^2-b^2}$ $(a>0,\ b>0)$, где b в несколько раз меньше a (и потому $\sqrt{a^2-b^2}\approx\sqrt{a^2}=a$; $\approx$ — знак приближенного равенства), при вычислениях можно преобразовать так:

$$Q=\frac{(a-\sqrt{a^2-b^2})(a+\sqrt{a^2-b^2})}{a+\sqrt{a^2-b^2}}=\frac{b^2}{a+\sqrt{a^2-b^2}},$$

после чего оно уже не будет содержать этих опасных разностей.

10. Умножение и деление приближенных чисел. Общие замечания. Начнем с примера. Пусть надо найти площадь S прямоугольника со сторонами $a=5{,}2$ *см* и $b=43{,}1$ *см*. Было бы неправильно дать такой ответ: $S=5{,}2\cdot 43{,}1=224{,}12$ *см*^{2}.

Действительно, на самом деле a заключено между 5,1 и 5,3 *см*, а b — между 43,0 и 43,2 *см* и потому площадь заключена между

$$S_1=5{,}1\cdot 43{,}0=219{,}3\ см^2 \quad и \quad S_2=5{,}3\cdot 43{,}2=228{,}96\ см^2,$$

т. е. в найденном значении S все цифры после второй сомнительные и могут только ввести в заблуждение. Ответ следует дать такой: $S=2{,}2\cdot 10^2$ *см*^{2}.

Заметим, кстати, что по тому образцу, как мы вычислили S_1 и S_2, и в других примерах можно дать двусторонние оценки для ответа.

Итак, мы видим, что при умножении двух чисел с двумя и тремя верными знаками в ответе следует оставить два верных знака. Аналогичное правило справедливо в общем случае, а также при делении приближенных чисел: в ответе число верных знаков надо взять равным наименьшему (худшему) числу верных знаков в сомножителях (или в делимом и в делителе, если рассматривается частное). Дело в том, что, как мы покажем в п. IX.11, *при умножении или делении приближенных чисел предельные относительные погрешности складываются*, а число верных знаков говорит примерно о том же, о чем и предельная относительная погрешность, т. е. об относительной точности.

В приведенном примере с вычислением S предельная относительная погрешность у b была значительно меньше, чем у a, а потому $\delta_S=\delta_a+\delta_b\approx\delta_a$, т. е. и число верных знаков у S такое же, как у a.

Если множители даны с разным числом верных знаков, то перед умножением следует произвести округление, оставив один лишний знак, который надо отбросить после выполнения действий. Если множители заданы с одинаковым числом верных знаков, но этих

множителей много, например более четырех, то верных знаков в произведении следует взять на один меньше.

Таким образом, например, при вычислении количества тепла, выделяемого электрическим током, по формуле $Q=0,24I^2Rt$ в ответе не может получиться более двух верных знаков, так как коэффициент 0,24 имеет лишь два верных знака; при этом нет смысла брать I, R и t более чем с тремя верными знаками (да и то третий знак, если берется, является лишь запасным). Если Q требуется с большей точностью, то надо прежде всего уточнить коэффициент.

Отметим, что совершенно точные множители не влияют на выбор числа верных знаков в произведении: например, в формуле для длины окружности $L=2\pi r$ коэффициент 2 является совершенно точным (он может быть записан в виде 2,0 или 2,00 и т. п.), так что точность, с которой можно вычислить L, зависит только от числа верных знаков, с которыми взято π и определено r.

Приведем пример на применение всех этих правил. Пусть $D=11,3^2\cdot 5,4+$ $+0,381\cdot 9,1+7,43\cdot 21,1$. Для выяснения, насколько велики слагаемые, вычислим их, произведя округление всех чисел до одного верного знака. Получаем 500, 3,6 и 140. Значит, сумма будет содержать несколько сотен, а поскольку в первом, самом большом слагаемом один из множителей (5,4) дан только с двумя верными знаками, то и весь ответ получится с двумя верными знаками. Согласно правилу лишнего знака будем проводить вычисления с точностью до единиц, а потом ответ округлим до десятков. Получится $D=690+3+157=850$, т. е. $D=8,5\cdot 10^2$

Вычисления с лишними цифрами были бы не только бесплодными, но даже вредными, дающими иллюзию точности, когда ее на самом деле нет.

При выборе степени точности приближенных величин, над которыми надо производить те или иные вычисления, руководствуются *принципом равной точности*, согласно которому все эти выбираемые степени точности должны быть согласованы друг с другом и ни одна не должна быть чрезмерной или недостаточной.

Поясним этот принцип на примерах. Пусть мы вычисляем площадь прямоугольника по формуле $S=ab$. Тогда, если a измерено или вычислено, например, с тремя верными знаками, то и b следует взять с тремя верными знаками, так как четвертый знак у b все равно будет излишним, а если b взять только с двумя верными знаками, то пропадет труд, затраченный на нахождение третьего знака у a. Таким образом, в произведении всегда выгодно множители (во всяком случае те, нахождение которых связано с теми или иными трудностями) брать с одним и тем же числом верных знаков. Аналогично в сумме надо брать слагаемые с одним и тем же числом знаков после запятой.

Приведем еще один пример. Пусть вычисляется выражение $M=ab+cd$, причем грубо известно, что $a\approx 30$, $b\approx 6$, $c\approx 0,1$ $d\approx 40$. Пусть a взято с тремя верными знаками; с каким числом знаков следует взять b, c и d? Ясно, что b надо взять также с тремя верными знаками, руководствуясь числом знаков у a. Далее $ab\approx 180$, $cd\approx 4$, т. е. чтобы M было известно с тремя верными знаками (а точнее оно не может получиться при выбранном a), c и d достаточно взять лишь с одним верным знаком. Если это нетрудно, то точность b, c и d или некоторых из этих величин следует повысить на один знак, но этот знак будет запасной.

В практике приближенных вычислений мы часто сталкиваемся с задачей, до некоторой степени обратной к только что рассмотренным. Бывает так, что степень точности результата вычислений задается заранее из каких-либо

дополнительных соображений, а требуется определить необходимую степень точности величин, участвующих в этих вычислениях (эти величины могут, например, получаться из опыта, так что речь идет о необходимой степени точности эксперимента), и степень точности вычислений. Решение этой обратной задачи опирается на приведенные правила приближенных вычислений. Пусть, например, вычисляется полная поверхность кругового цилиндра по формуле $S=\pi\left(DH+\frac{D^2}{2}\right)$, причем грубо известно, что $D\approx 20$ *см*, $H\approx 2$ *см*. Тогда $S\approx 700$ *см*^{2} (проверьте!). Рассуждая аналогично предыдущему абзацу, увидим, что если, например, ответ желательно иметь с тремя верными знаками, т. е. с точностью до 1 *см*^{2}, то π и D надо взять с тремя верными знаками, а H — с двумя, т. е. измерение D и H проводить с точностью до 1 *мм*. Вычисление следует проводить с одним запасным знаком, да и π следует взять с одним запасным знаком, тогда как запасной знак при измерении D и H в данном случае потребовал бы повышения класса точности измерения.

Правила расчета степени точности для более сложных формул будут указаны в п IV.10 и IX.11.

§ 3. Функции и графики

11. Функциональная зависимость. Часто бывает, что в одном и том же рассмотрении участвует одновременно несколько переменных величин, взаимосвязанных друг с другом таким образом, что *изменение одних величин сказывается на значениях других*. Тогда говорят, что между рассматриваемыми величинами имеется *функциональная зависимость*. Например, при изменении условий, в которых содержится какая-либо определенная порция газа, функциональная зависимость будет между объемом V, температурой T и давлением p этого газа, так как эти величины взаимосвязаны. Функциональная зависимость имеется между площадью круга и длиной его радиуса, между пройденным путем и временем в процессе движения и т. п.

Обычно среди функционально зависимых между собой величин можно указать некоторые величины (*независимые переменные*), значения которых могут выбираться более или менее произвольно, тогда как значения остальных величин (*зависимых переменных*) определяются значениями первых. Например, при рассмотрении связи между площадью S круга и длиной R его радиуса эту длину естественно принять за независимую переменную, так как ее значения можно задавать произвольно; при этом площадь, определяемая по формуле $S=\pi R^2$, будет зависимой переменной. При указанном выше рассмотрении порции газа за независимые переменные можно взять V и T; давление p будет тогда зависимой переменной.

Закон (правило), по которому значениям независимых переменных отвечают (соответствуют) значения рассматриваемой зависимой переменной, называется функцией. Таким образом, каждый раз, когда нам дан такой закон соответствия, мы можем сказать: вот функция. Функция — одно из важнейших математических понятий.

Впрочем, слово «функция» употребляется и в ином смысле. Именно, часто независимые переменные называются также *аргументами*, а зависимая переменная — *функцией* от этих аргументов. Обычно такое двоякое употребление слова «функция» не приводит к ошибкам.

Следует отметить, что если между величинами имеется функциональная зависимость, то часто выбор того, какие из этих величин считать независимыми, а какие — зависимыми, является довольно условным. Так, в приведенном примере с порцией газа за независимые переменные можно было бы принять T и p, а V — за зависимую переменную; нетрудно привести схему опыта, в котором бы T и p задавались, а объем V находился. Выбор того, какие переменные более естественно или более удобно принять за независимые, иногда довольно важен. (Как известно, кавалер де Монбасон на вопрос: «Что к чему привешено: хвост к собаке или собака к хвосту?» ответствовал: «Как, сударыня, приключится; ибо всякую собаку никому за хвост, как и за шею, приподнять невозбранно»; однако ясно, что эти способы неравноценны.)

Функции могут быть от одного аргумента (как в примере площади круга) или от двух и более аргументов. В первых главах нашего курса мы будем рассматривать почти исключительно функции от одного аргумента.

Заметим, что для того, чтобы некоторая величина y могла рассматриваться как функция от независимой переменной x, нет надобности, чтобы между изменениями этих величин существовала глубокая причинная связь. Достаточно только, чтобы существовал определенный закон, по которому значениям x отвечали бы значения y, этот закон может быть нам и неизвестен. Например, температуру θ в какой-либо точке можно считать функцией времени t, так как ясно, что значениям t отвечают определенные значения θ, хотя, конечно, изменение θ объясняется не просто течением времени, но рядом глубоких физических причин.

12. Обозначения. Если величина y является функцией от величины x, то обычно пишут $y=f(x)$ (читается: «*игрек* есть *эф* от *икс*»), где f, начальная буква латинского слова functio, — знак функции. *Частные значения* этой функции получаются, если аргументу x придавать частные (конкретные) значения.

Пусть, например, $y=f(x)$ имеет такой вид: $y=x^2$. Тогда при $x=2$ будет $y=4$, при $x=-0{,}6$ будет $y=0{,}36$ и т. п. Это можно записать так: $f(2)=4$, $f(-0{,}6)=0{,}36$ и т. д., или $y|_{x=2}=4$, $y|_{x=-0{,}6}=0{,}36$ и т. д.; здесь вертикальная черта — знак подстановки, который означает, что вместо аргумента x подставлено какое-либо его значение.

Запись вида $y=f(x)$ применяется, если конкретное выражение функции слишком громоздкое или даже нам не известно, а также для формулировки правил и свойств, общих для всех или многих

конкретных функций (как, например, в алгебре формула $(a+b)^3 = a^3+3a^2b+3ab^2+b^3$ приводится не для конкретных чисел, а для букв, вместо которых можно подставить любые конкретные числа).

Если одновременно рассматривается несколько различных функций, то, кроме f, приходится применять другие буквы F, φ, Φ и т. п. или применять *индексы* (значки): f_1, f_2 и т п. Однако в разных рассмотрениях одной и той же буквой f можно обозначить различные функции, как в алгебре в одной и той же задаче буквой a нельзя обозначить различные величины, но в другой задаче та же буква a может означать что-либо другое. Если же разные величины связаны одинаковой зависимостью, то можно применять один и тот же знак функции, так как f означает закон зависимости одной величины от другой. Например, если $y=x^3$, $z=u^5$, $v=t^3$, то можно написать $y=f(x)$, $z=\varphi(u)$, $v=f(t)$; в данном случае знак f означает возведение аргумента в третью степень, а знак φ — в пятую.

Аналогично обозначаются функции от нескольких аргументов. Пусть, например, $z=x^2-x2^y$, x и y — независимые переменные, z — зависимая; тогда можно написать $z=f(x, y)$, запятая в данном случае является признаком функции от двух аргументов. В этом случае частные значения находятся так:

$$f(2, 1) \text{ (т. е. } z\big|_{\substack{x=2\\y=1}}) = 2^2-2\cdot 2^1=0; \quad f(1, 2)=1^2-1\cdot 2^2=-3 \text{ и т. п.}$$

Ко всем этим обозначениям надо привыкнуть и свободно с ними манипулировать Приведем несколько примеров таких манипуляций. Пусть рассматриваются две функции $y=f(x)=x^2-3x$ и $z=\varphi(x)=2x+1$, а a — постоянное число. Тогда

$f(a)=a^2-3a$ (значение первой функции при $x=a$);

$\varphi(a^2)=2a^2+1$ (значение второй функции при $x=a^2$),

$f(x^2)=(x^2)^2-3x^2=x^4-3x^2$ [значение y, если вместо аргумента подставлено x^2, получается новая функция от x, которую можно обозначить, например, через $F(x)$],

$[f(x)]^2=(x^2-3x)^2=x^4-6x^3+9x^2$ (еще новая функция x);

$\varphi(x+a)=2(x+a)+1=2x+2a+1$ (новая функция x);

$f(x)\varphi(x)=(x^2-3x)(2x+1)=2x^3-5x^2-3x$;

$f(\varphi(x))=[\varphi(x)]^2-3\varphi(x)=(2x+1)^2-3(2x+1)=4x^2-2x-2$;

$\varphi(f(x))=2f(x)+1=2(x^2-3x)+1=2x^2-6x+1$;

$f(x+s)=(x+s)^2-3(x+s)=x^2+2xs+s^2-3x-3s$ [функция двух независимых переменных, которую можно обозначить через $\Phi(x, s)$] и т п.

В разобранных примерах мы сталкиваемся, в частности, с образованием «функции от функции» или, как говорят, с образованием *сложной функции*. Обычно сложная функция получается следующим образом. Пусть переменная y зависит от переменной u, которая в свою очередь зависит от переменной x, т. е. $y=f(u)$, $u=\varphi(x)$. Тогда при изменении x будет меняться u, а потому будет меняться и y. Значит, y является функцией x, $y=f(\varphi(x))$, которая и назы-

вается сложной функцией; переменная u в данном случае называется *промежуточной*. Может быть и несколько промежуточных переменных.

Если хотят только отметить, что y является функцией от x, но не производить подобные манипуляции, то пишут просто $y=y(x)$; так, в примерах п 11 $S=S(R)$, $p=p(V, T)$, $V=V(T, p)$.

13. Способы задания функций. Чтобы функцию, т е. зависимость одной величины от другой, можно было изучить, она должна быть как-то задана. Имеется несколько способов задания функции.

Аналитический способ (при помощи формулы) чаще всего применяется в математике. В этом способе явно указываются математические действия, которые надо совершить над независимой переменной, чтобы получить значение функции. Например, формула $y=x^2-2x$ означает, что для того, чтобы получить значение функции y, нужно значение аргумента возвести в квадрат и из результата вычесть удвоенное значение этого аргумента.

Аналитический способ компактен (формула занимает мало места), легко воспроизводи́м (формулу легко переписать) и наиболее приспособлен к выполнению над функциями математических действий — алгебраических (сложение, умножение и т. п.), действий высшей математики (дифференцирование, интегрирование и т. п.) и других. Однако он не всегда нагляден (не всегда виден характер зависимости функции от аргумента) и для вычисления значений функции, если они требуются, необходимо произвести ряд выкладок, не всегда простых; да и *далеко не все функции возможно и целесообразно записать в виде формулы*.

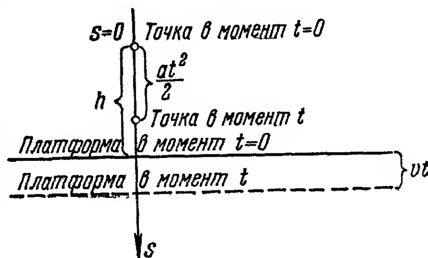


Рис. 6.

Функция иногда задается несколькими формулами, действующими на различных участках изменения аргумента. Пусть, например, материальная точка падает без начальной скорости на платформу, которая перемещается равномерно вниз со скоростью v и отстоит в начальный момент $t=0$ от точки на расстоянии h. Тогда путь s, пройденный точкой, зависит от времени t, $s=f(t)$, причем зависимость эта определяется формулами (рис. 6)

$$s=f(t)=\begin{cases} \dfrac{at^2}{2} & (0\leqslant t\leqslant t^*), \\ h+vt & (t^*\leqslant t<\infty), \end{cases}$$

где t^* — момент столкновения точки с платформой, определяемый из уравнения $\frac{at^2}{2}=h+vt$.

В табличном способе задания функции ее численные значения задаются с помощью таблицы при определенных дискретных численных значениях аргумента. Таблица имеет такой вид:

$$y = f(x)$$

x	x_1	x_2	x_3	...	x_N
y	$y_1 = f(x_1)$	$y_2 = f(x_2)$	$y_3 = f(x_3)$	...	$y_N = f(x_N)$

(2)

Каждая из разностей $x_2 - x_1$, $x_3 - x_2$, ... называется *шагом таблицы*. Наиболее удобны таблицы с *постоянным шагом*, для которых аргумент x принимает значения a, $a+h$, $a+2h$, ... (h — шаг). Хорошо известные таблицы логарифмов, тригонометрических функций и т. п. — это примеры табличного способа задания функции; в этих таблицах для экономии места значения функции выписываются не в единый ряд, как (2), а строчками, как слова в книге. Имеются таблицы и многих других важных функций. Таблицы часто получаются в результате обработки эксперимента, когда задаются отдельными значениями одной величины и измеряют значения другой и т. д.

Большим удобством табличного способа является то, что значения функции уже вычислены, так что ими можно немедленно пользоваться. Однако могут понадобиться значения функции при значениях аргумента, которых нет в таблице; тогда приходится производить дополнительные вычисления — *интерполяцию* (для промежуточных значений аргумента) или *экстраполяцию* (для значений аргумента, лежащих за пределами таблицы), что иногда приводит к неверным результатам. Таблицы часто занимают большой объем, составляются с затратой большого труда; правда, в последнее время с развитием вычислительной техники таблицы стали вычисляться значительно быстрее Недостатком этого способа является также то, что он плохо приспособлен к выполнению математических действий: ведь каждое такое действие требует составления новой таблицы, что является трудоемким делом и не всегда может быть выполнено достаточно точно.

Третьим основным способом задания функции является *графический способ* (с помощью графика). Этот способ очень нагляден, так как по графику легко детально проследить за характером изменения функции. Кроме того, по графику можно быстро находить значения функции с небольшой точностью (два-три верных знака), правда, только в изображенном диапазоне изменения аргумента. Однако составление графика с достаточной точностью требует затраты определенного труда, а точность, с которой получаются значения функции из графика, может оказаться недостаточной. Отметим,

что графики может вычерчивать самопишущий прибор в процессе эксперимента.

За последние годы широко распространился и стал одним из основных четвертый способ задания функции — с помощью *указания программы* для вычисления ее значений на цифровой вычислительной машине. С этим способом мы познакомимся в п. XIX 6.

Все эти способы задания функции как бы дополняют друг друга, так что часто возникает задача о переходе от одного способа к другому — о построении графика, о составлении таблицы (так называемое *табулирование*), о подборе формулы. В нашем курсе мы столкнемся с такими задачами.

Встречаются также и иные способы задания функций. Например, закон, по которому значения функции соответствуют значениям аргумента, иногда формулируется словесно: так, ежемесячный членский взнос может быть функцией заработной платы.

Первое определение понятия функции, близкое к современному, дал в 1718 г. швейцарский математик И. Бéрнулли (1667—1748); но в XVIII в. функция обычно отождествлялась с аналитической формулой. Современное общее понятие функции как закона зависимости впервые возникло у Эйлера в 1755 г., но утвердилось только в XIX в.

14. Графики функций. Графики служат для геометрического изображения функций. Напомним методику построения графиков функций, известную из курса средней школы. Пусть величина y является функцией величины x, т. е. $y=f(x)$. Для построения графика на плоскости выбираются две числовые оси: обычно ось переменной x проходит слева направо и называется *осью абсцисс*, а ось переменной y проходит перпендикулярно к оси x и называется *осью ординат*. Начало отсчета на каждой из осей часто выбирается в точке их взаимного пересечения (рис. 7). После этого придают аргументу всевозможные значения, находят соответствующие значения $y=f(x)$ и строят точки графика.

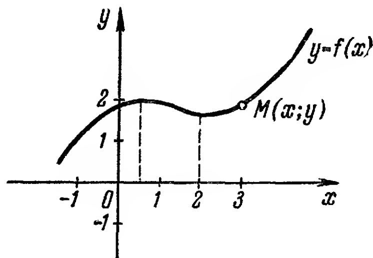


Рис. 7.

На рис. 7 показана произвольная «текущая» точка M графика, имеющая координаты x, y. Практически мы можем построить не очень большое число точек графика, после чего соединяем их линией; теоретически же необходимо представлять себе, как будто переменная x пробегает *всю* область своего изменения; тогда текущая точка M пробежит весь график. На рис. 7 показан пример графика. Из него видно, что в данном случае при возрастании аргумента x значение функции сначала возрастает; это продолжается

примерно до значения $x=0,5$, после чего функция убывает, сравнительно медленно; начиная же примерно с $x=2$, функция вновь возрастает, причем все быстрее и быстрее.

Единицы масштаба и начала отсчета на каждой из осей выбираются так, чтобы лучше всего передать ход изменения функции на наиболее интересных интервалах изменения аргумента и функции.

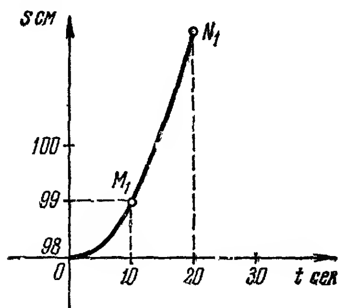


Рис. 8.

Рассмотрим, например, график равноускоренного движения, протекающего по закону

$$s=98+0,01t^2 \qquad (t \geqslant 0), \qquad (3)$$

где t выражено в *сек*, а s — в *см*. В этом случае возможно выбрать шкалы на обеих осях так, как показано на рис. 8 Ясно, что изменение начала отсчета на оси аргумента (или оси функции) влечет за собой перенос графика как целого параллельно оси аргумента (или соответственно оси функции). Изменение масштаба на какой-либо из осей в несколько раз влечет за собой растяжение во столько же раз графика от другой оси (или сжатие к ней); например, на рис. 9 показан график той же функции (3) после изменения масштаба по оси t Новый график получается из старого растяжением от оси s параллельно оси t.

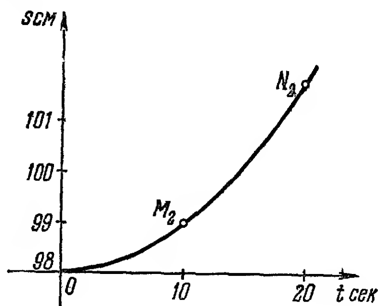


Рис. 9.

Чтобы наилучшим образом передать поведение рассматриваемой функции, иногда применяются на осях неравномерные шкалы, о которых говорилось в п. 4.

В дальнейшем мы будем всегда считать, если не оговорено противное, что переменные (аргументы и функция) — безразмерные. В этом случае в теории проще всего считать, как мы и будем делать, *что единицы масштаба по обеим осям одинаковые, а отсчет ведется от точки их пересечения*, которая называется *началом координат*. О том, как влияет изменение масштаба или начала отсчета, упоминалось раньше.

15. Область определения функции. *Областью определения функции* называется совокупность значений независимой переменной, при которых эта функция определена, т. е. область изменения независимой переменной (п. 5). Обычно эта переменная является непрерывной, и тогда, как было указано в п. 5, эта область определения состоит из одного или нескольких интервалов.

В некоторых случаях область определения функции выясняется из физического или геометрического смысла этой функции. Например, если рассматривается зависимость $S=\pi R^2$ площади круга от длины его радиуса, то областью определения этой функции будет интервал $0<R<\infty$, так как по геометрическому смыслу R может принимать именно такие значения. Если рассматривается зависимость плотности ρ атмосферы над данной точкой земной поверхности от высоты h над уровнем моря, то областью определения этой функции будет интервал $h_0 \leqslant h \leqslant H$, где h_0 — высота земной поверхности, а H — условная высота, принимаемая за границу атмосферы, и т. д. Если функция задана просто формулой, то областью определения служит совокупность значений аргумента, при которых формула дает определенное вещественное (действительное) значение функции. (Мы пока будем рассматривать только *вещественные функции от вещественного аргумента*, т. е. функции, у которых зависимая и независимая переменные принимают лишь вещественные значения.) Например, если $y=x^3$, то x может принимать любые значения, т. е. областью определения служит вся числовая ось $-\infty<x<\infty$. Если $y=\sqrt{x^2-2}$, то при вычислении y встретится препятствие в извлечении корня, если окажется, что $x^2-2<0$; значит, должно быть $x^2-2 \geqslant 0$, т. е. $x^2 \geqslant 2$, а это будет при $x \leqslant -\sqrt{2}$ или $x \geqslant \sqrt{2}$, т. е. область определения в данном случае состоит из двух интервалов: $-\infty<x \leqslant -\sqrt{2}$ и $\sqrt{2} \leqslant x<\infty$ (на рис. 10 эта область заштрихована). При нахождении области определения в аналогичных случаях надо выяснить, что́ может препятствовать получению значения функции, после чего выписывать неравенства (как в последнем примере $x^2-2 \geqslant 0$), гарантирующие возможность этого получения. Тогда задача сведется к решению этих неравенств.

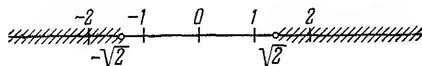


Рис. 10.

Если независимая переменная дискретна, то область определения функции состоит из дискретных (отдельных) точек. Например, если $f(x)=x!=1\cdot 2\,\ldots\,x$, то x может принимать только значения 1, 2, 3, ... Если, как в этом примере, дискретный аргумент принимает лишь целые значения, то обычно его обозначают не x, а буквами n, m, k и т. п., а вместо $f(1)$, $f(2)$, ..., $f(n)$, ... пишут a_1, a_2, ..., a_n, ... и говорят, что дана *последовательность*; например, последовательностью служит геометрическая прогрессия

$$a_1=a, \quad a_2=aq, \quad a_3=aq^2, \ \ldots, \ a_n=aq^{n-1}, \ \ldots$$

и т. п. График функции от дискретного аргумента не является линией, а состоит из дискретных точек (рис. 11).

Область изменения самой функции называется иначе *множеством значений этой функции*. Например, для функции $y=x^2$ областью

определения служит интервал $-\infty < x < \infty$, а множеством значений — интервал $0 \leqslant y < \infty$, так как в данном случае y принимает только такие значения.

Выяснение области определения функции важно для построения ее графика, так как эта область — это та часть оси абсцисс, над или под которой пройдет график; точнее говоря, это — проекция графика на ось абсцисс. На рис. 12 показаны три простых графика; области определения этих функций заштрихованы. Ясно, что если область определения состоит из нескольких частей, то и график состоит из нескольких кусков.

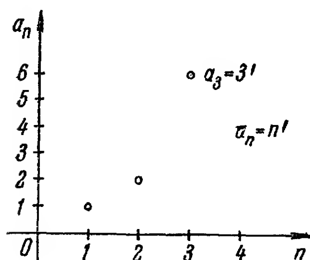


Рис. 11.

16. Характеристики поведения функции. Надо научиться свободно характеризовать различные качества функции аналогично тому, как мы характеризуем качества людей: спокойный, блондин и т. п.

Всюду, где не оговорено противоположное, мы будем считать исследуемые функции *однозначными*, т. е. считать, что каждому значению независимой переменной из ее области изменения отвечает

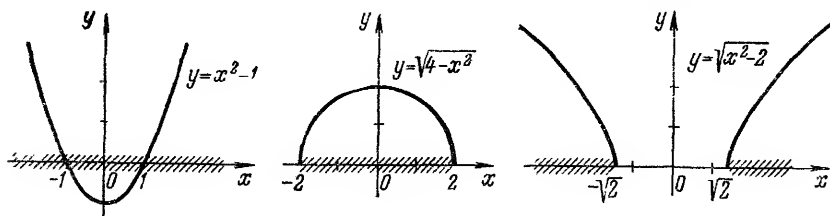


Рис. 12.

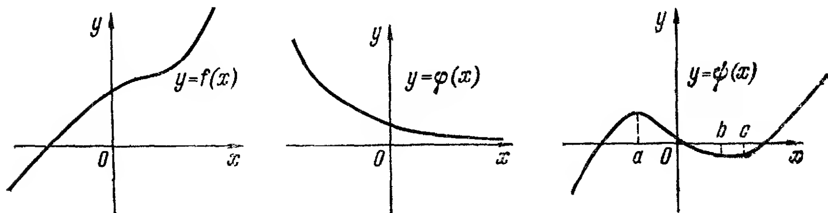


Рис. 13.

одно вполне определенное значение функции. О *многозначных*, т. е. неоднозначных функциях будет сказано в п. 20.

Функция называется *возрастающей* (соответственно *убывающей*), если при росте аргумента значения функции возрастают (соответственно убывают). Как возрастающие, так и убывающие функции на-

зываются *монотонными*. Если функция не является монотонной, то на оси аргумента можно указать *интервалы монотонности*, на которых функция монотонна, иногда они чередуются с *интервалами постоянства* функции. Так, на рис. 13 показаны графики возрастающей функции $f(x)$, убывающей функции $\varphi(x)$ и немонотонной функции $\psi(x)$; последняя функция имеет интервал возрастания $-\infty < x \leqslant a$, интервал убывания $a \leqslant x \leqslant b$, интервал постоянства $b \leqslant x \leqslant c$ и интервал возрастания $c \leqslant x < \infty$.

Условие возрастания функции $f(x)$ можно записать так: из $x_1 < x_2$ всегда следует, что $f(x_1) < f(x_2)$. Это дает возможность производить одинаковые действия над обеими частями неравенства: например, зная, что $y = x^3$ — возрастающая функция, мы получаем, что из неравенства $a < b$ всегда вытекает неравенство $a^3 < b^3$ и обратно. Если функция $f(x)$ не является монотонной, то такие действия можно производить на интервале ее возрастания, на интервале убывания функции $f(x)$ из $x_1 < x_2$ вытекает, что $f(x_1) > f(x_2)$. Например, функция $y = x^2$ — убывающая при $-\infty < x \leqslant 0$ и возрастающая при $0 \leqslant x < \infty$; значит, из $a < b$ при $b \leqslant 0$ вытекает $a^2 > b^2$, а при $a \geqslant 0$ вытекает $a^2 < b^2$.

Функция называется *непрерывной*, если при постепенном (непрерывном) изменении аргумента значения функции меняются также

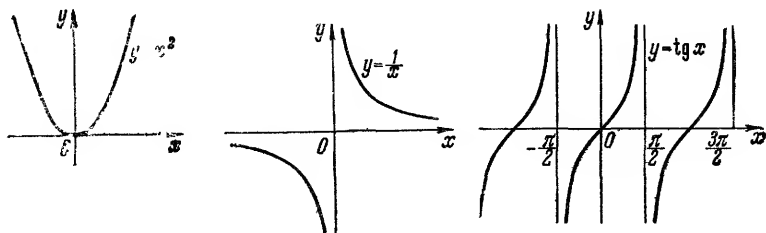


Рис. 14.

постепенно, без скачков. В противном случае функция называется *разрывной*, а значения аргумента, при которых непрерывность (постепенность) изменения функции нарушается, называются *точками разрыва* функции. (Более подробное разъяснение этих понятий будет дано в § III.4.) Так (рис. 14), функция $y = x^2$ непрерывна на всей оси x; функция $y = \frac{1}{x}$ имеет одну точку разрыва $x = 0$ (при приближении аргумента к значению $x = 0$ значения функции уходят в бесконечность), а при остальных x функция непрерывна; функция $y = \operatorname{tg} x$ имеет бесконечное число точек разрыва $x = \pm\frac{1}{2}\pi, \ \pm\frac{3}{2}\pi, \ \pm\frac{5}{2}\pi, \ldots$

Если функция определена с обеих сторон от точки разрыва, то график этой функции также разрывен и состоит из двух или большего числа частей (кусков; см., например, рис. 14).

Функция $y = \sin x$ является примером периодической функции. Именно (рис. 15), поведение этой функции на интервалах

$$\ldots; \quad -4\pi \leqslant x \leqslant -2\pi; \quad -2\pi \leqslant x \leqslant 0; \quad 0 \leqslant x \leqslant 2\pi;$$
$$2\pi \leqslant x \leqslant 4x, \ \ldots$$

совершенно одинаковое; более точно: $\sin(x+2\pi) \equiv \sin x$. Здесь $\equiv$ *знак тождественного равенства*; его пишут, когда хотят подчеркнуть,

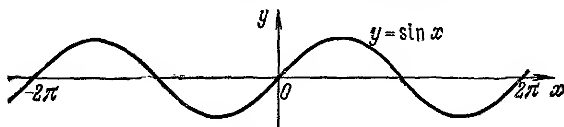


Рис. 15.

что некоторое равенство является тождеством (хотя можно писать и обычный знак равенства).

Число 2π называется периодом функции $y = \sin x$. Конечно, имеют место также тождества

$$\sin(x+4\pi) \equiv \sin x, \quad \sin(x+6\pi) \equiv \sin x,$$
$$\sin[x+(-2\pi)] \equiv \sin x \ \text{и т. д.},$$

однако за период принимается положительное значение, причем обычно

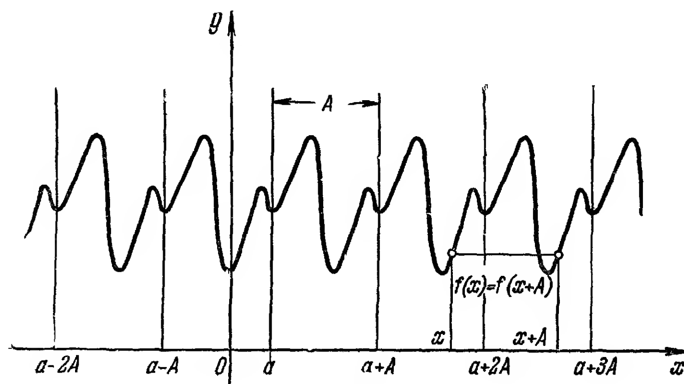


Рис. 16.

наименьшее возможное, т. е. не 4π или 6π, а именно 2π; впрочем, это значение иногда называют *наименьшим периодом*.

В общем случае функция $y = f(x)$ называется *периодической с периодом* $A > 0$, если имеет место тождество $f(x+A) \equiv f(x)$. Поведение такой функции на каждом из интервалов

$$\ldots; \quad a-2A \leqslant x \leqslant a-A; \quad a-A \leqslant x \leqslant a; \quad a \leqslant x \leqslant a+A;$$
$$a+A \leqslant x \leqslant a+2A; \ \ldots$$

(где a — произвольное выбранное число) совершенно одинаковое (рис. 16), так что достаточно рассматривать функцию на одном из таких отрезков. (На рис. 16 показано также равенство $f(x+A)=f(x)$ для одного из значений x.)

Функция $f(x)$ называется *четной*, если она не меняется при изменении знака у аргумента; другими словами, если $f(-x)\equiv f(x)$. Примерами четных функций служат $y=x^2$, $y=x^6$, $y=\cos x$ и т. д. Из рис. 17 видно, что график четной функции симметричен относительно оси ординат. Функция $f(x)$ называется *нечетной*, если при

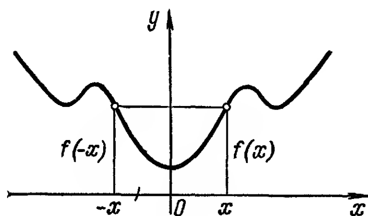


Рис. 17.

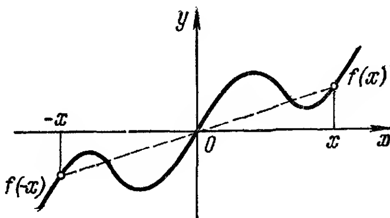


Рис. 18.

изменении знака у аргумента она умножается на -1, т. е. если $f(-x)\equiv -f(x)$. Примерами могут служить $y=x$, $y=x^5$, $y=\sin x$ и т. д. Из рис. 18 видно, что график нечетной функции симметричен относительно начала координат. Отметим, что функция может не быть ни четной, ни нечетной; например, $y=1+\sin x$, $y=1-x$, $y=2^x$, $y=\lg x$ и т. д.

17. Алгебраическая классификация функций. Функции, заданные единой формулой (см. п. 13), классифицируются в зависимости от характера алгебраических действий, которые надо совершить над аргументом, чтобы получить значение функций. Если применяются только сложение, вычитание и умножение, а также возведение в целую положительную степень, которое является частным случаем умножения, то функция называется целой рациональной, или многочленом; при образовании многочлена могут применяться произвольные постоянные коэффициенты. Примеры многочленов: $y=x^3-2x+3$; $y=x^2$; $y=3$; $y=-\frac{x}{\pi}+\sqrt{2}\,x^3$; $y=a^4x^2-2$ и т. д. С другой стороны, функции $y=x^{-5}$, $y=x^3+2\sqrt{x}$ не являются многочленами в смысле приведенного определения. Каждый многочлен имеет *степень*, которая определяется старшей из участвующих степеней независимой переменной: так, степени приведенных многочленов равны 3, 2, 0, 3, 2.

Более широкий класс составляют *рациональные* функции, в которых допускается также и деление; при этом, если рациональная функция не является целой, то она называется дробной рациональной

функцией. Пример:

$$y=\frac{x^2-\frac{1}{x-1}}{x\sqrt{2}-3}-\frac{ax-\sqrt[3]{3}}{x^{-2}+1}.$$

После приведения к общим знаменателям по правилам элементарной алгебры *всякую дробную рациональную функцию можно представить в виде отношения двух многочленов.*

Еще более широкий класс составляют *алгебраические* функции, в которых допускается также и извлечение корня; при этом, если алгебраическая функция не является рациональной, она называется *иррациональной.* Пример иррациональной функции: $y=x^2-\frac{1}{x}+\sqrt{x^2-1}$.

Неалгебраические функции называются *трансцендéнтными.* Примеры трансцендентных функций: $y=\sin x$, $y=x^2+\operatorname{tg} x$, $y=2^x$, $y=\lg x$ и т. д.; отметим, что две последние функции являются трансцендентными, несмотря на то, что в школе они по традиции рассматриваются в курсе алгебры.

Все эти определения автоматически переносятся на функции нескольких переменных. Единственным новым моментом является определение степени многочлена при помощи сложения показателей степеней аргументов в одночленах.

Так, функция $f(x, y)=x^4y-x^4y^2+x$ является многочленом шестой степени от x и y; если же, например, в этой функции считать y зафиксированным, то она будет многочленом четвертой степени от x.

Для любого числа переменных многочлен первой степени называется *линейной* функцией, многочлен второй степени — *квадратичной* и т. д.

18. Элементарные функции. Перечислим основные функции, применяемые в элементарной математике:

$y=x^a$ (при постоянном a) — *степеннáя* функция;

$y=a^x$ (при постоянном a) — *показательная* функция, она же называется *экспоненциáльной* функцией или просто *экспонéнтой*;

$y=\log_a x$ (при постоянном a) — *логарифмическая* функция;

$y=\sin x$, $\cos x$, $\operatorname{tg} x$, $\operatorname{ctg} x$ — *тригонометрические* функции;

$y=\arcsin x$, $\arccos x$ и т. д. — *обратные тригонометрические* функции.

Элементарными функциями называются все функции, которые можно составить из основных элементарных функций с помощью алгебраических действий (с применением любых коэффициентов) и образования сложных функций (см. п. 12). Так, все алгебраические функции являются элементарными. Но элементарны и очень многие трансцендентные функции, например $x+\lg\sin x$, $2^{\lg\operatorname{tg}x+\sin x}$ и т. д. (возможны очень большие нагромождения).

Элементарные функции составляют значительную часть функций, которые рассматриваются в общем курсе высшей математики. Примером неэлементарной функции может служить, скажем, $y=x!$. В специальных отделах математики неэлементарные функции используются очень широко; многие из этих функций очень хорошо изучены и систематически применяются, так что в настоящее время выделение «элементарных» функций в значительной мере устарело.

19. Преобразования графиков. Часто бывает, что известны графики каких-либо функций, а требуется построить графики других функций, так или иначе выражающихся через первые. Мы приведем несколько примеров таких преобразований графиков.

Пусть дан график функции $y=f(x)$ и требуется построить графики функций $z=f(x)+a$ и $u=f(x+b)$ (a и b — постоянные), причем величины z и u будем откладывать по той же оси, что и y (рис. 19). Тогда при любом x будет $z=y+a$, т. е. график функции $z(x)$ получается из графика функции $y(x)$ при помощи переноса вдоль оси y на a в положительном направлении (см. рис. 19, где каждый из вертикальных отрезков имеет длину a). Что касается графика функции $u(x)$, то часто по ошибке говорят, что он получается из графика $y(x)$ переносом на b вдоль оси x в положительном направлении. На самом деле направление переноса получается прямо противоположное. Действительно, чтобы $u=y$, надо в выражении для u взять аргумент на b меньше, чем в выражении для y, так как тогда $u=f[(x-b)+b]=f(x)=y$. Поэтому график функции $u(x)$ получается из графика функции $y(x)$ переносом на b вдоль оси x в отрицательном направлении. Конечно, если $a<0$ и $b<0$, то перенос будет в противоположном направлении, но этого не надо особо оговаривать, так как всегда подразумевается, что перенос на «—3» вверх — это все равно, что перенос на «3» вниз.

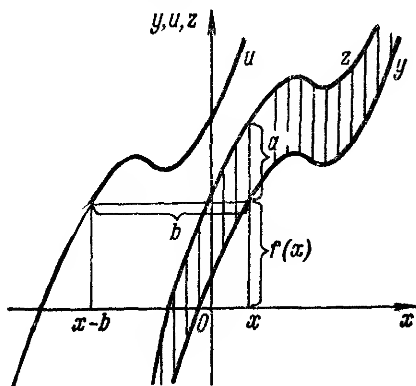


Рис. 19.

Подобным же образом строятся графики функций $v=kf(x)$ и $w=f(kx)$ (рис. 20). График функции $v(x)$ получается из графика функции $y(x)$ путем равномерного растяжения от оси x в k раз, так как точки первого графика имеют при тех же абсциссах ординаты в k раз больше, чем у второго. График же функции $w(x)$ получается из графика функции $y(x)$ равномерным сжатием к оси

y в k раз, так как $w\left(\frac{x}{k}\right)=f\left(k\frac{x}{k}\right)=f(x)=y(x)$. Конечно, если $0<k<1$, то вместо растяжения будет сжатие, и наоборот (всегда подразумевается, что растяжение в $\frac{1}{3}$ раза — это сжатие в три раза). Если $k<0$, то к указанным преобразованиям добавится еще зеркальное отражение графика от оси x для функции $v(x)$ или от оси y для функции $w(x)$.

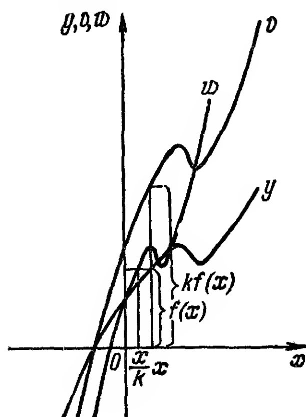


Рис. 20.

Комбинируя эти результаты, видим, что график функции $y=kf(mx+b)+a$ получается из графика $y=f(x)$ путем последовательных преобразований: переноса [который даст график $y=f(x+b)$], сжатия $[y=f(mx+b)]$, растяжения $[y=kf(mx+b)]$ и еще одного переноса.

Аналогичных результатов можно достичь, сохраняя график, но производя соответствующие манипуляции над осями координат. Например, вместо переноса графика вправо можно перенести оси координат влево, другими словами, переместить влево начало отсчета на оси x; вместо растяжения графика от оси x в k раз можно уменьшить в k раз единицу масштаба по оси y.

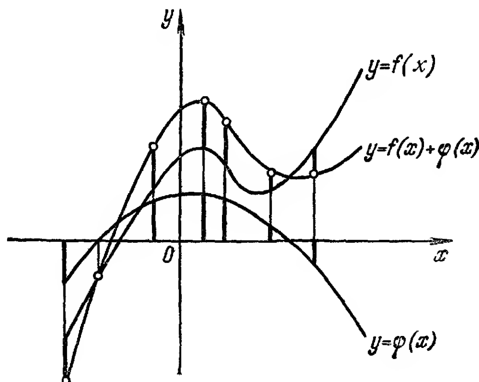


Рис. 21.

Над функциями, заданными графически, можно производить арифметические действия. Например, на рис. 21 показано графическое сложение двух функций: графики функций $f(x)$ и $\varphi(x)$ даны, а график их суммы строится, причем расположенные друг над другом равные отрезки выделены жирными линиями.

Покажем в заключение (рис. 22) графическое построение сложной функции $z=\varphi[f(x)]$, если известны графики каждой из функций $z=\varphi(y)$ и $y=f(x)$. Эти графики удобно расположить так, как показано на рисунке, после чего, задаваясь значениями x, переносить отрезок $A'B'$ в положение AB; при этом точка B опишет требуемый график сложной функции.

20. Неявные функции. *Неявной функцией* называется функция, определенная из неразрешенного уравнения, связывающего аргумент и функцию. Разрешая это уравнение, мы получаем ту же функцию, но уже заданную в явной форме. Так, равенства $x-y^3+2=0$ и $y=\sqrt[3]{x+2}$ равносильны; они определяют одну и ту же функцию $y(x)$, но первое равенство определяет ее в неявной форме, как неявную функцию, а второе — в явной. Часто бывает, что разрешить уравнение относительно функции невозможно или нецелесообразно; в этом случае уравнение так и оставляют неразрешенным, в общей форме (после переноса всех членов в левую часть)

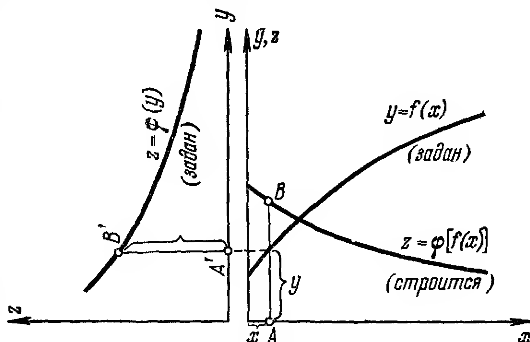


Рис. 22.

$$F(x, y)=0. \tag{4}$$

Этого не нужно бояться, так как позже мы узнаем ряд приемов, приспособленных к изучению функций, заданных в неявной форме.

Если в уравнении (4), определяющем неявную функцию $y(x)$, задавать значения независимой переменной x, то для нахождения соответствующего значения y надо решать уравнение. Как известно, если в уравнение подставить его решение, то получится тождество. Поэтому можно сказать также, что неявная функция $y=y(x)$, определенная уравнением (4), — это такая функция, которая, будучи подставлена в уравнение (4), обращает его в тождество (проверьте это на приведенном выше примере).

Уравнение (4) при заданном x может иметь более одного решения. Тогда функция $y(x)$ будет многозначной, т. е. при заданном значении аргумента принимает более одного значения. Например, рассматривая неявную функцию $y(x)$, определенную из уравнения

$$x-y^2=0, \tag{5}$$

мы получаем при любом заданном $x>0$ два значения y: $y=\sqrt{x}$ и $y=-\sqrt{x}$; *само значение радикала всегда подразумевается взятым в арифметическом смысле.* Рассмотрение многозначных функций неудобно и его стараются избежать, разбивая такую функцию на

однозначные ветви, отвечающие тому или иному значению функции. Так, в нашем примере двузначная функция $y=\pm\sqrt{x}$, определенная из уравнения (5), имеет две однозначные ветви: $(y)_1=\sqrt{x}$ и $(y)_2=-\sqrt{x}$.

Каждая ветвь неявной функции представляет собой однозначную функцию и потому имеет график обычного вида. Все эти ветви составляют обычно (исключения будут указаны в п. II.8) единую линию, которая и является графиком функции, определенной уравнением (4). Так, в нашем примере уравнение (5) можно переписать в виде $x=y^2$, откуда ясно, что графиком служит обычная «школьная» парабола, но необычно расположенная, так как оси x и y

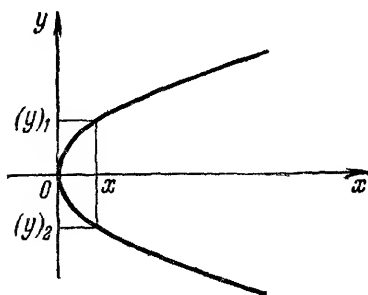


Рис. 23.

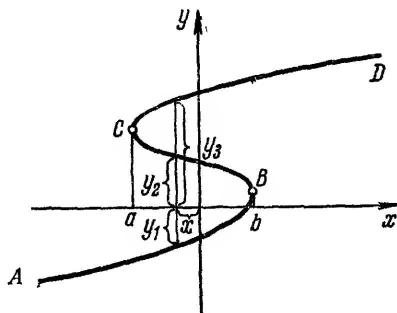


Рис. 24.

поменялись ролями по сравнению со «стандартным» уравнением $y=x^2$ (рис. 23). Каждая из однозначных ветвей изображается половиной параболы, первая — верхней, вторая — нижней.

График неявной функции может иметь, например, вид, изображенный на рис. 24. Здесь видно, что при $x<a$ и при $x>b$ функция $y(x)$ является однозначной, а при $a<x<b$ — трехзначной; при разделении значений на ветви естественно считать дугу AB графиком первой ветви, дугу BC — графиком второй и CD — графиком третьей ветви.

В связи с рассмотрением многозначных функций заметим, что возможны такие функции, для которых каждому значению независимой переменной отвечает целый интервал значений функции. Например, такое соотношение будет между ростом и возможным весом человека. Такие функции рассматриваются обычно в теории вероятностей (п. XVIII. 16) и в других главах нам встречаться не будут.

21. Обратные функции. Пусть рассматривалась функция

$$y=f(x). \qquad (6)$$

Будем придавать y различные значения и находить соответствующие значения x, т. е. примем бывшую зависимую переменную за

аргумент, а бывшую независимую — за функцию. Полученная функция (зависимость) $x(y)$ называется *обратной* по отношению к исходной функции $y(x)$. Она задается тем же равенством (6), в котором, однако, надо рассматривать y как независимую переменную, а x — как зависимую. Но в п. 12 мы уже обращали внимание на то, что при рассмотрении одной и той же функции можно по-разному обозначать переменные. Поэтому если мы захотели бы для обратной к (6) функции обозначить, как обычно, независимую переменную через x, а зависимую — через y, то надо просто подставить в (6) x вместо y, а y вместо x, т. е. равенство, определяющее обратную функцию, надо переписать в виде

$$x = f(y). \tag{7}$$

Таким образом, обратная функция оказывается заданной в неявной форме и поэтому (см. п. 20) оказывается, вообще говоря, многозначной. Легко указать условие однозначности обратной функции — им служит

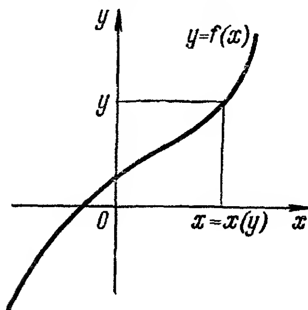


Рис. 25.

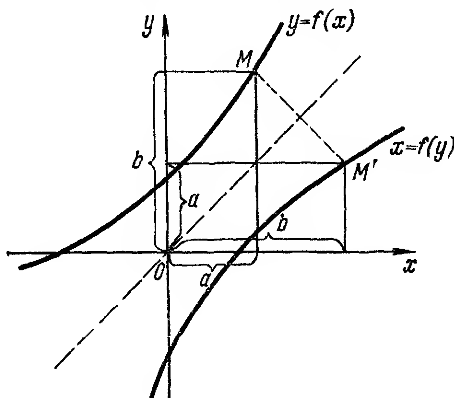


Рис. 26.

монотонность исходной функции, так как тогда, задаваясь значениями y, мы каждый раз будем получать единственное значение $x = x(y)$ (рис. 25).

Примеры: Обратной к $y = x^3$ служит функция, определенная из равенства $x = y^3$, т. е. $y = \sqrt[3]{x}$; обратной к $y = x^2$ служит двузначная функция $y = \pm\sqrt{x}$.

Равенства (6) и (7) получаются в результате простой перестановки величин x и y, т. е. в результате перемены их ролей. Поэтому из рис. 26 видно, что график обратной функции получается из графика исходной функции с помощью зеркального отражения последнего относительно биссектрисы угла между осями координат, указанной на рис. 26 пунктиром. (Обе точки M и M' на рис. 26 отвечают одному и тому же равенству вида $b = f(a)$.)

В заключение отметим, что если функция $x(y)$ обратна по отношению к функции $y(x)$, то, наоборот, вторая обратна по отношению к первой; эти функции являются взаимно обратными.

§ 4. Обзор простейших функций

Многие из функций, которые мы здесь рассмотрим, известны из курса средней школы. Они собраны вместе из-за их большого значения.

22. Линейная функция. Линейная функция (см. конец п. 17) имеет общий вид

$$y = ax + b, \tag{8}$$

где a и b — постоянные коэффициенты.

Графиком линейной функции служит прямая линия (рис. 27). Коэффициент a называется *угловым коэффициентом* этой прямой; чем $|a|$ больше, т. е. чем a больше по абсолютному значению, тем прямая идет круче. Если

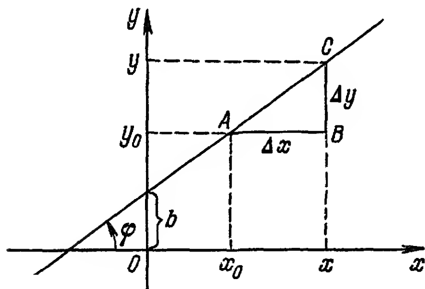


Рис. 27.

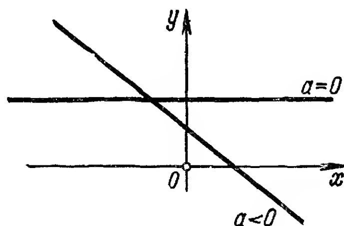


Рис. 28.

аргумент изменился от некоторого значения x_0 до значения x, получив *приращение**) Δx; а функция получила соответствующее приращение Δy, то из равенств $y_0 = ax_0 + b$, $y = ax + b$ следует $y - y_0 = a(x - x_0)$, т. е. $\Delta y = a\Delta x$ и

$$\frac{\Delta y}{\Delta x} = a. \tag{9}$$

Итак, для линейной функции отношение приращения функции к приращению аргумента постоянно и равно угловому коэффициенту графика; *приращение линейной функции прямо пропорционально приращению аргумента.*

На рис. 27 изображен случай, когда $a > 0$. Если $a < 0$, то прямая проходит направо вниз (рис. 28). Если $a = 0$, то прямая парал-

*) Δ — обычный знак для приращения; Δx надо рассматривать как единый символ и ни в коем случае не как произведение Δ на x. «Приращение» понимается в алгебраическом смысле, оно может быть и отрицательным.

лельна оси x; в этом случае функция постоянна, т. е. получается график константы.

На свойстве приращения линейной функции основана *линейная интерполяция*, которая применяется уже в школьной практике и состоит в следующем. Пусть значения некоторой функции $y=f(x)$, график которой изображен на рис. 29 пунктиром, известны при $x=x_0$, и $x=x_0+h$,

$$f(x_0)=y_0, \quad f(x_0+h)=y_1,$$

но неизвестны при промежуточных значениях x. Тогда мы приближенно заменяем данную функцию линейной, принимающей те же значения при $x=x_0$ и $x=x_0+h$, т. е. заменяем дугу AE отрезком прямой. Из подобия треугольников ABC и ADE получаем тогда

$$\frac{y-y_0}{x-x_0}=\frac{y_1-y_0}{h}, \quad \text{т. е.}$$

$$y=y_0+\frac{y_1-y_0}{h}(x-x_0).$$

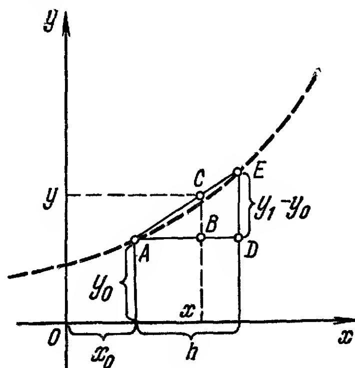


Рис. 29.

Такая замена возможна, если функция $f(x)$ на рассматриваемом интервале мало отличается от линейной. Она широко применяется, в частности, для таблиц с достаточно малым шагом, когда последовательные значения функции мало отличаются друг от друга. Более точные методы интерполяции, учитывающие нелинейность функции, будут освещены в пп. V.6—8. Аналогично осуществляется *линейная экстраполяция* (см. п. 13).

Из формулы (9) и рис. 27 видно, что $a=\operatorname{tg}\varphi$, т. е. *угловой коэффициент прямой равен тангенсу угла, под которым она наклонена к оси абсцисс.*

Если величины x и y размерные, то и угловой коэффициент размерный. Из формулы (8) видно, что $[b]=[y]$, $[ax]=[y]$, откуда $[a]=\frac{[y]}{[x]}$ (аналогично выясняется размерность коэффициентов в других формулах). Легко выяснить геометрический смысл углового коэффициента: если единице величины x отвечает l_x единиц длины по оси x, а единице величины y отвечает l_y единиц длины по оси y (l_x и l_y—это *масштабные коэффициенты*), то треугольник ABC на рис. 27 имеет длины сторон $l_x\Delta x$ и $l_y\Delta y$, откуда

$$\operatorname{tg}\varphi=\frac{l_y\Delta y}{l_x\Delta x} \quad \text{и} \quad a=\frac{\Delta y}{\Delta x}=\frac{l_x}{l_y}\operatorname{tg}\varphi, \tag{10}$$

т. е. угловой коэффициент пропорционален указанному тангенсу.

23. Квадратичная функция. Квадратичная функция в общем виде такова:

$$y = ax^2 + bx + c.$$

В курсе средней школы показывается, что *графиком квадратичной функции служит парабола*. В наиболее простом случае, когда $a=1$, $b=0$, $c=0$, т. е. $y=x^2$, график показан на рис. 30. Тогда функция будет четной, т. е. ось y для нее служит осью симметрии (*ось параболы*). Точка пересечения параболы с ее осью называется *вершиной параболы*; на рис. 30 эта вершина расположена в начале координат.

В общем случае, при любых a, b, c, парабола получится в результате равномерного растяжения и параллельного переноса из той

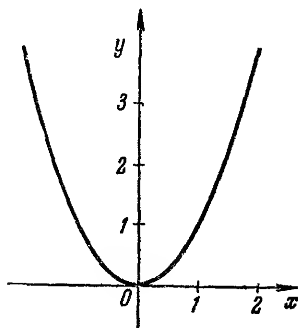


Рис. 30

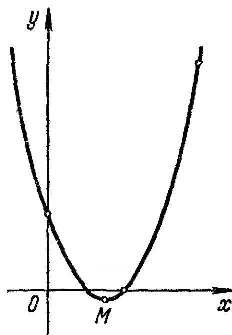


Рис. 31.

параболы, которая изображена на рис. 30. При этом выяснить положение вершины можно по методу *дополнения до полного квадрата*, который мы поясним на числовом примере. Пусть $y = 2x^2 - 3x + 1$ *). Тогда совершаем следующие простые преобразования:

$$y = 2\left(x^2 - \frac{3}{2}x + \frac{1}{2}\right) = 2\left[\left(x - \frac{3}{4}\right)^2 - \left(\frac{3}{4}\right)^2 + \frac{1}{2}\right] =$$

$$= 2\left(x - \frac{3}{4}\right)^2 - \frac{1}{8}.$$

Таким образом (см. п. 19), требуемый график получается из параболы, изображенной на рис. 30, в результате переноса вправо на $\frac{3}{4}$, равномерного растяжения от оси x в два раза и последующего переноса вниз на $\frac{1}{8}$. Полученный график изображен на рис. 31;

*) На практике обычно встречаются не такие трехчлены, а трехчлены, например, вида $y = 2{,}17x^2 - 3{,}21x + 0{,}84$, коэффициенты которых известны лишь приближенно. Однако если разобрать случай простых коэффициентов, то потом нетрудно перейти и к более сложному случаю. Это относится и к дальнейшим подобным примерам.

для более точного его построения следует придать x несколько значений и найти соответствующие значения y, после чего построить соответствующие точки на графике (например, при $x=0$, 1 и 2 получается $y=1$, 0 и 3; соответствующие точки на графике отмечены). Вершина полученной параболы находится в точке M с координатами $x=\frac{3}{4}$, $y=-\frac{1}{8}$. Эта парабола более узкая, чем изображенная на рис. 30 (при той же единице масштаба).

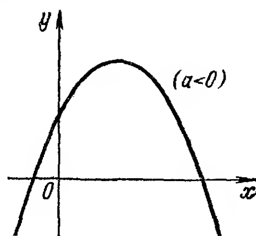


Рис. 32.

Вообще, чем больше $|a|$, тем парабола у́же. Если $a<0$, то парабола уходит вниз (рис. 32), а если $a=0$, то квадратичная функция превращается в линейную.

24. Степенна́я функция. Степенная функция имеет вид

$$y=x^n.$$

Если $0<x<1$, то чем больше n, тем значение функции меньше; если же $x>1$, то чем больше n, тем значение функции больше. Соответствующие графики при $n=1$, 2, 3, 4 изображены на рис. 33. При построении графиков в сторону $x<0$ надо учесть, что при четном n функция получается четной, а при нечетном n — нечетной. Обратим, в частности, внимание на график функции $y=x^3$ (*кубическая парабола*). При $x<0$ график *выпуклый кверху* (вогнутый книзу), т. е. лежит под касательной, проведенной в любой его точке. При $x>0$ график *выпуклый книзу*. В начале координат выпуклость в одну сторону сменяется выпуклостью в другую сторону; здесь касательной к графику служит ось x, однако в точке касания O график переходит с одной стороны касательной на другую. Такие точки называются *точками перегиба* данной кривой линии. Таким образом, кубическая парабола имеет одну точку

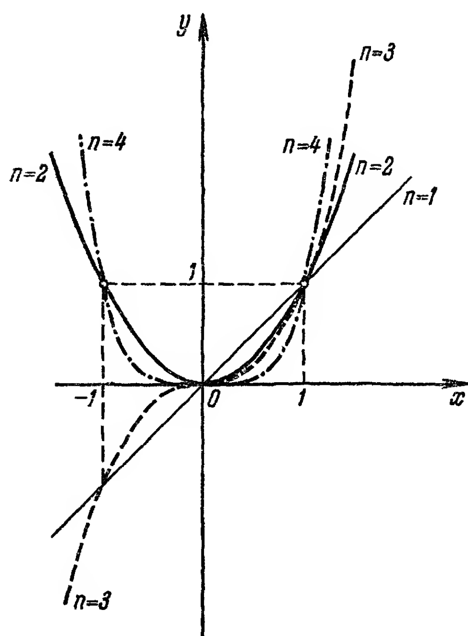


Рис. 33.

перегиба; из широко известных линий точками перегиба обладает, например, синусоида.

При n нецелых графики располагаются между соответствующими графиками для целых n. Однако в этом случае при построении графика для отрицательных x надо соблюдать осторожность, так как отрицательное число в нецелой степени может дать мнимое значение; в этом случае график для $x<0$ не строится.

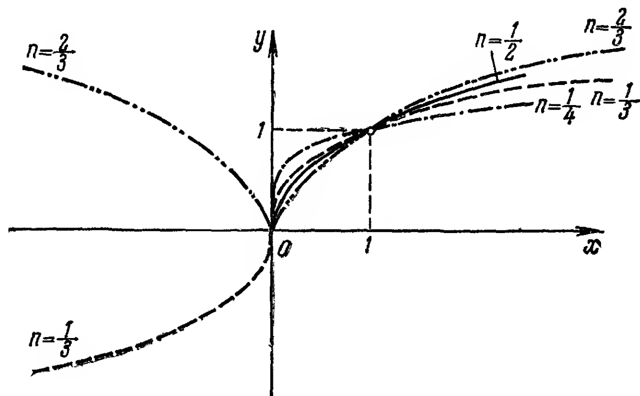


Рис. 34.

Рассмотрим случай $0<n<1$. Пусть, например, $n=\frac{1}{2}$, т. е. $y=x^{\frac{1}{2}}=\sqrt{x}$. Тогда, как было показано в п. 20, графиком будет служить верхняя половина обычной (квадратной) параболы с осью, расположенной по оси x (рис. 34). На рис. 34 изображены графики степенных функций при некоторых других дробных n. Если дробь, представляющая n, имеет нечетный знаменатель, то график существует не только при $x>0$, но и при $x<0$, так как из отрицательных чисел возможно извлекать корень с нечетным показателем. Обратим, в частности, внимание на график функции $y=x^{2/3}$ (*полукубическая парабола*), воспроизведенный на рис. 35. Этот график, подходя к началу координат, отходит от него вспять, имея там острие, так называемую *точку возврата*. В дальнейшем мы познакомимся с другими линиями, имеющими точки возврата.

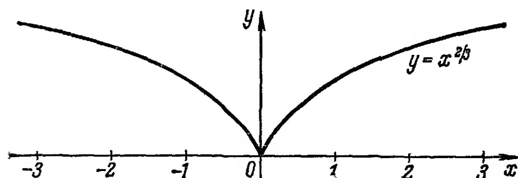


Рис. 35.

Рассмотрим, наконец, случай отрицательного $n=-m$. Тогда $y=\frac{1}{x^m}$ и потому при весьма малых $|x|$ получаются весьма большие $|y|$ и наоборот. Соответствующие графики показаны на рис. 36 при $x>0$; чтó будет при $x<0$, мы предоставим разобрать читателю. Все эти графики при удалении в бесконечность вытягиваются вдоль координатных осей, безгранично к ним приближаясь. Если кривая и

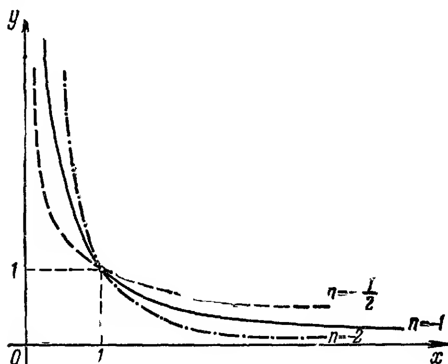


Рис. 36.

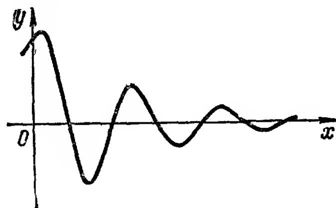


Рис. 37.

прямая расположены таким образом друг относительно друга, то прямая называется *асимптóтой* этой кривой; таким образом, каждый из указанных графиков имеет по две асимптоты, которыми служат оси координат.

Не следует думать, что и в других случаях кривая не может пересекать свою асимптоту. Так, при рассмотрении затухающих колебаний получается график вида, изображенного на рис. 37. Здесь ось x также служит асимптотой графика.

25. Дробно-линейная функция. Дробно-линейная функция представляет собой отношение двух линейных функций и потому имеет общий вид

$$y=\frac{ax+b}{cx+d}. \qquad (11)$$

В самом простом случае, когда $a=d=0$, если обозначить $\frac{b}{c}=k$, получим $y=\frac{k}{x}$, т. е. обратную пропорциональную зависимость. Соответствующий график, как известно из средней школы, называется *гиперболой*. На рис. 38 этот график изображен в двух случаях: когда $k>0$ и когда $k<0$. Будучи графиком нечетной функции, гипербола имеет центр симметрии, на рис. 38 им служит начало координат; она обладает двумя асимптотами, на рис. 38 ими являются оси координат. В п. II.13 мы проверим, что гипербола имеет две оси симметрии, на рис. 38 ими служат биссектрисы углов между осями координат.

В общем случае *графиком дробно-линейной функции* также *служит гипербола*, но параллельно перенесенная по сравнению с поло-

жением, изображенным на рис. 38. Покажем это на числовом примере. Пусть $y=(2x+3)/(3x-5)$. Проведем простые преобразования:

$$y=\frac{2\left(x+\frac{3}{2}\right)}{3\left(x-\frac{5}{3}\right)}=\frac{2}{3}\frac{x-\frac{5}{3}+\frac{5}{3}+\frac{3}{2}}{x-\frac{5}{3}}=$$

$$=\frac{2}{3}\frac{\left(x-\frac{5}{3}\right)+\frac{19}{6}}{x-\frac{5}{3}}=\frac{2}{3}\left[1+\frac{\frac{19}{6}}{x-\frac{5}{3}}\right]=\frac{\frac{19}{9}}{x-\frac{5}{3}}+\frac{2}{3}.$$

Таким образом (см. п. 19), требуемый график получается из графика $y=\frac{19/9}{x}$ переносом направо на $\frac{5}{3}$ и вверх на $\frac{2}{3}$.

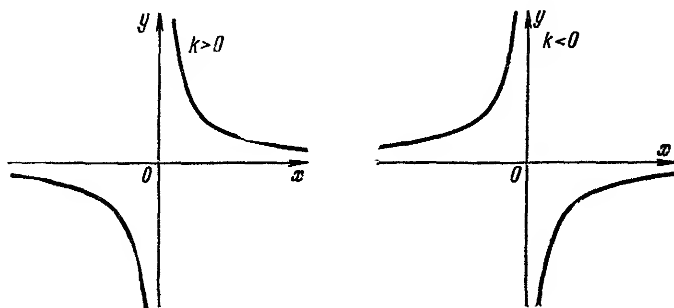


Рис. 38.

Получится гипербола с центром симметрии в точке $x=\frac{5}{3}$; $y=\frac{2}{3}$ (рис. 39).

Дробно-линейная функция общего вида (11) имеет точку разрыва при $x=-\frac{d}{c}$, где знаменатель обращается в нуль. Поэтому и получилось (см. п. 16), что ее график состоит из двух кусков.

26. Логарифмическая функция. Логарифмическая функция — это функция вида

$$y=\log_a x. \tag{12}$$

Она определена только при $x>0$, причем рассматривается при основаниях $a>0$ $(a\neq 1)$. Графики логарифмических функций при различных основаниях показаны на рис. 40. Они не имеют ни оси симметрии, ни центра симметрии, но обладают асимптотой, которой служит ось y. *Все логарифмические функции пропорциональны друг другу*, так как, прологарифмировав равенство $a^{\log_a x}=x$ по основанию b, получаем

$$\log_b x=\log_a x\cdot\log_b a=k\log_a x \qquad \left(k=\log_b a=\frac{1}{\log_a b}\right). \tag{13}$$

Таким образом, все графики рис. 40 получаются из одного путем равномерного растяжения от оси x или сжатия к этой оси. Для дальнейшего имеет значение тот угол, под которым эти графики пересекают ось x; конечно, имеется в виду угол между осью x и касательной к графику в точке пересечения, так как *углом между двумя линиями в точке их пересечения называется угол между касательными к ним в этой точке.* При указанном растяжении графиков касательная поворачивается, причем для очень больших a она

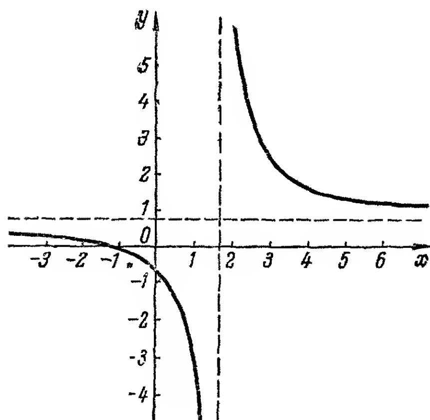


Рис 39.

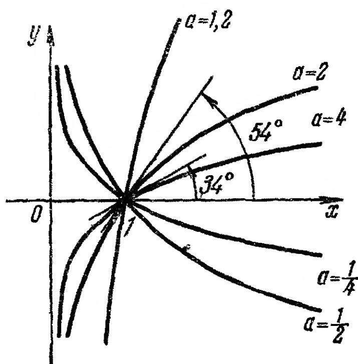


Рис. 40.

наклонена весьма полого, а при a, близких к 1,— весьма круто. *При некотором значении a угол пересечения графика логарифмической функции* (12) *с осью x равен* 45°; *это значение обозначается буквой e* и играет в высшей математике, как мы увидим в дальнейшем, чрезвычайно большую роль.

На рис. 40 видно, что при $a=2$ рассматриваемый угол пересечения больше 45°, а при $a=4$ — меньше; значит, e заключено между 2 и 4. Точные подсчеты, о которых будет сказано в п. IV.16, показывают, что $e=2{,}71828$, с точностью до 10^{-5}. Обозначение числа e ввел Эйлер.

Логарифм по основанию e называется *натуральным логарифмом* и обозначается $\ln x=\log_e x$. График натурального логарифма показан на рис. 41. Логарифм при любом другом основании можно выразить через натуральные логарифмы согласно формуле (13):

$$\log_a x=\frac{\ln x}{\ln a}. \qquad (14)$$

Например, формулы для перехода от десятичных логарифмов к натуральным и обратно таковы:

$$\lg x=0{,}4343\ln x, \quad \ln x=2{,}303\lg x,$$

где коэффициент пропорциональности указан с четырьмя верными знаками.

Помимо натуральных логарифмов, в математике широко применяются десятичные (при численных расчетах) и двоичные (в теории информации и некоторых других современных отделах).

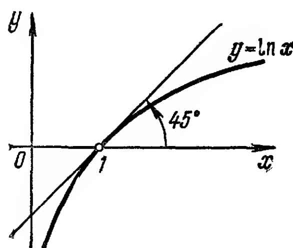


Рис. 41.

27. Показательная функция. Показательной функцией называется функция

$$y = a^x. \qquad (15)$$

Она определена при всех x, причем рассматривается только при основаниях $a > 0$, так как для $a < 0$ при возведении в нецелую степень результат может получиться мнимым. Равенство (15) получится, если формулу (12) разрешить относительно x, что даст $x = a^y$, а затем переставить x и y. Таким образом (см. п. 21), показательная и логарифмическая функции обратны друг другу.

Поэтому графики показательных функций, показанные при различных основаниях на рис. 42, получаются из соответствующих графиков рис. 40 логарифмических функций с помощью зеркального отражения относительно биссектрисы угла между осями координат. Если $a > 1$, то показательная функция является возрастающей, причем тем быстрее, чем больше a. Если $0 < a < 1$, то показательная функция убывающая.

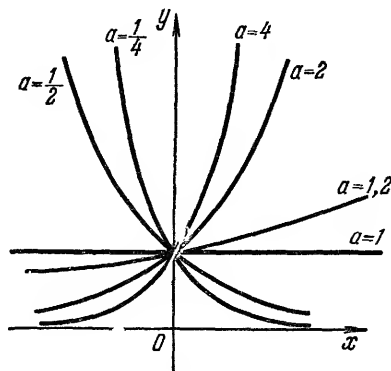


Рис. 42.

Обычно за основание показательной функции принимается число e; в этом случае показательная (экспоненциальная) функция имеет специальное обозначение $y = e^x = \exp x$.

Показательную функцию с другим основанием можно привести к основанию e: так как по определению логарифма $a = e^{\ln a}$, то $a^x = (e^{\ln a})^x = e^{kx}$ $(k = \ln a)$.

28. Гиперболические функции. *Гиперболическими синусом, косинусом и тангенсом* называются функции

$$\operatorname{sh} x = \frac{e^x - e^{-x}}{2}; \quad \operatorname{ch} x = \frac{e^x + e^{-x}}{2}; \quad \operatorname{th} x = \frac{\operatorname{sh} x}{\operatorname{ch} x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Эти названия могут сначала показаться несколько странными, и подлинный их смысл (что здесь от синуса? что здесь от гиперболы?) будет ясен лишь в п. VIII.4 и XIV.8. Установим некоторые формулы, связывающие эти функции друг с другом, для чего возведем первые два равенства в квадрат:

$$\operatorname{sh}^2 x = \frac{e^{2x} - 2 + e^{-2x}}{4}; \quad \operatorname{ch}^2 x = \frac{e^{2x} + 2 + e^{-2x}}{4}.$$

Вычитая и складывая эти две формулы, получим

$$\operatorname{ch}^2 x - \operatorname{sh}^2 x = 1; \quad \operatorname{ch}^2 x + \operatorname{sh}^2 x = \frac{e^{2x} + e^{-2x}}{2} = \operatorname{ch} 2x.$$

Уже эти полученные формулы указывают на значительную аналогию

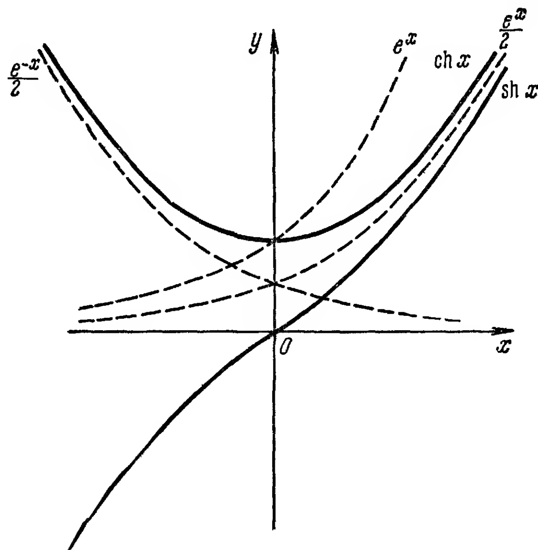


Рис. 43.

между гиперболическими функциями и тригонометрическими. Мы предоставим читателю доказать формулы

$$\operatorname{sh} 2x = 2 \operatorname{sh} x \operatorname{ch} x; \quad \operatorname{sh}(a+b) = \operatorname{sh} a \operatorname{ch} b + \operatorname{ch} a \operatorname{sh} b; \quad 1 - \operatorname{th}^2 x = \frac{1}{\operatorname{ch}^2 x}$$

и другие аналогичные формулы по своему усмотрению. Заметим, что $\operatorname{sh} 0 = 0$, а $\operatorname{ch} 0 = 1$; функции $\operatorname{sh} x$ и $\operatorname{th} x$ нечетные, а функция $\operatorname{ch} x$ четная: например,

$$\operatorname{sh}(-x) = \frac{e^{(-x)} - e^{-(-x)}}{2} = \frac{e^{-x} - e^{x}}{2} = -\frac{e^{x} - e^{-x}}{2} = -\operatorname{sh} x.$$

Построение графиков $\operatorname{sh} x$ и $\operatorname{ch} x$ показано на рис. 43, а график $\operatorname{th} x$, который можно построить по точкам, пользуясь первыми

двумя графиками, показан на рис. 44. Ясно, что гиперболические функции не обладают важнейшим свойством тригонометрических функций — свойством периодичности. Кроме того, множество значений (см. п. 15) каждой гиперболической функции существенно отличается от множества значений соответствующей тригонометрической функции.

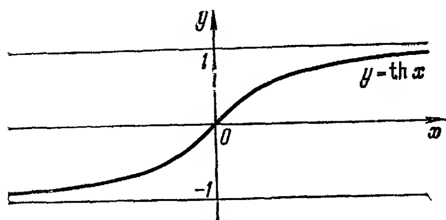


Рис. 44.

График функции $\operatorname{th} x$ имеет две асимптоты, так как, например, при больших x будет $e^{-x} \ll 1 \ll e^{x}$ (здесь $\ll$ — знак «*значительно меньше*») и потому $\operatorname{th} x \approx 1$.

Иногда рассматриваются *обратные гиперболические функции*; они обозначаются $\operatorname{arsh} x$, $\operatorname{arch} x$ и $\operatorname{arth} x$. Из рис. 43 и 44 видно, что первая и третья функции являются однозначными (ср. рис. 25), тогда как вторая — двузначная. Все эти функции можно выразить через логарифм. Действительно, пусть, например, $y = \operatorname{arsh} x$. Тогда по определению обратной функции

$$x = \operatorname{sh} y = \frac{e^{y} - e^{-y}}{2},$$

т. е.

$$e^{y} - e^{-y} - 2x = 0; \quad e^{2y} - 2xe^{y} - 1 = 0,$$

откуда $e^{y} = x \pm \sqrt{x^2+1}$. Так как левая часть положительна, то и правая должна быть положительна, т. е. перед радикалом можно взять только «+». Логарифмируя, получим

$$y = \operatorname{arsh} x = \ln(x + \sqrt{x^2+1}). \tag{16}$$

29. Тригонометрические функции. Периодическая с периодом 2π функция $y = \sin x$ хорошо известна из курса тригонометрии; ее

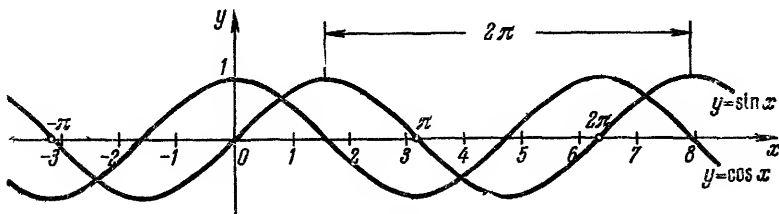


Рис. 45.

график (*синусоида*) показан на рис. 45. Эта функция нечетная, не имеет точек разрыва и ограничена (заключена между -1 и $+1$). Так как $\cos x = \sin(x + \pi/2)$, то график функции $\cos x$ — это та же синусоида, но сдвинутая на $\pi/2$ влево; он также показан на рис. 45. В приложениях часто синусоидальная, «*гармоническая*» зависимость появляется в виде

$$y = M \sin(\omega t + \alpha), \tag{17}$$

где независимая переменная t — время, постоянная $M>0$ называется *амплитудой*, $\omega>0$ *частотой* (круговой), сумма $\omega t+\alpha$ — *фазой*, постоянная α — *начальной фазой* (она получается из фазы при $t=0$). Легко выяснить, как влияют параметры M, ω и α на форму и расположение синусоиды (ср. п. 19). Амплитуда M увеличивает размах синусоиды от $-M$ до M, частота ω делает период вместо 2π равным $T=2\pi/\omega$, а из-за наличия начальной фазы синусоида сдвигается влево на α/ω, так как $\omega t+\alpha=\omega(t+\alpha/\omega)$, т. е. к аргументу прибавляется α/ω. Получившийся график показан на рис. 46.

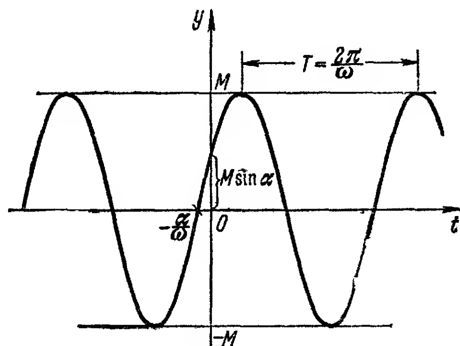


Рис. 46.

Функция вида (17) получается, в частности, при преобразовании выражения $A\cos\omega t+B\sin\omega t$. Так как правую часть (17) можно переписать в виде $M\sin\alpha\times\times\cos\omega t+M\cos\alpha\sin\omega t$, то для равенства

$$A\cos\omega t+B\sin\omega t\equiv\equiv M\sin(\omega t+\alpha)\qquad(18)$$

должно быть $A=M\sin\alpha$, $B=M\cos\alpha$. Отсюда легко найти M и α: $M=\sqrt{A^2+B^2}$, $\operatorname{tg}\alpha=\frac{A}{B}$; четверть, в которой нужно взять α, определяется знаками $\sin\alpha$ и $\cos\alpha$, т. е. знаками A и B.

Если независимой переменной служит не время, а геометрическая координата, то синусоидальную зависимость обычно записывают, взамен (17), в форме $y=M\sin(kx+\alpha)$. В этом случае k называется *волновым числом*, а $\lambda=2\pi/k$ — *длиной волны*.

Функция $y=\operatorname{tg}x$ имеет период π, так как $\operatorname{tg}(x+\pi)\equiv\operatorname{tg}x$ и точки разрыва при $x=\frac{\pi}{2}$; $\frac{\pi}{2}+\pi$; $\frac{\pi}{2}-\pi$; ... $\left(\text{в общем виде можно написать } x=\frac{\pi}{2}+k\pi, \text{ где } k=0, \pm1, \pm2, \ldots\right)$, где $\cos x=0$ и потому $\operatorname{tg}x=\pm\infty$. График этой функции (*тангенсоида*) показан на рис. 47; он состоит из бесконечного числа одинаковых кусков и имеет бесконечное число асимптот. На рис. 47 показан также график функции $y=\operatorname{ctg}x$. Так как

$$\operatorname{ctg}x=-\operatorname{tg}\left(x-\frac{\pi}{2}\right),$$

то линия получается та же, но иначе расположенная. (Как?)

Функция $y=\operatorname{Arcsin}x$ обратна по отношению к функции $y=\sin x$, поэтому график первой (рис. 48) получается из графика второй

путем зеркального отражения относительно биссектрисы угла между осями координат. Эта функция многозначная (точнее, бесконечнозначная) и потому (см. п. 20—21) обычно рассматривается ее *главная ветвь* (главное значение арксинуса), которая показана на рис. 48 более жирно; эта ветвь обозначается

$$y = \arcsin x, -\frac{\pi}{2} \leqslant \arcsin x \leqslant \frac{\pi}{2}$$

и представляет собой однозначную функцию. Остальные ветви специального названия и обозначения не имеют.

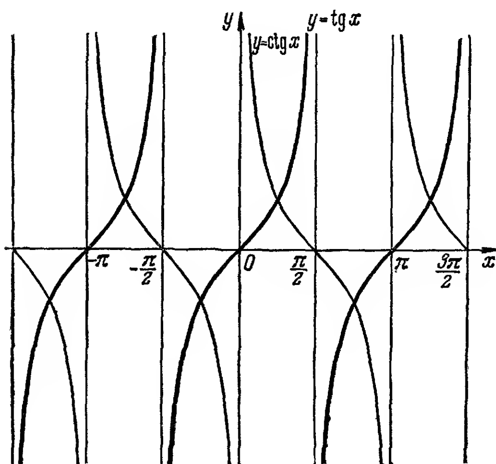


Рис. 47.

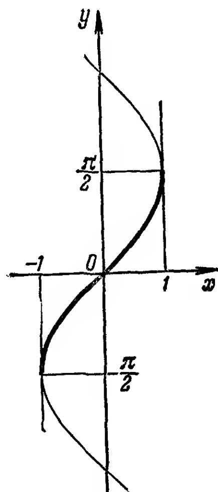


Рис. 48.

Аналогичное рассмотрение функций $y = \operatorname{Arccos} x$ и $y = \operatorname{Arctg} x$ мы предоставим читателю.

В заключение отметим, что значения $\arcsin x$ всегда будут браться отвлеченными (безразмерными), например

$$2^{\arcsin 1} = 2^{\frac{\pi}{2}} = 2^{1,57} = 2,97\ ^{*}).$$

Аналогичным образом $\sin x$ берется от отвлеченного значения x; при этом, конечно, синус числа x — это синус угла в x радиан. Например, $\sin 1 = \sin 57°18' = 0,8415$.

---

*) Конечно, два последних равенства приближенные. Если хотят это подчеркнуть, то пишут $2^{\frac{\pi}{2}} \approx 2^{1,57}$ и т. д. Впредь мы не будем делать таких оговорок.

30. Подбор эмпирической формулы. Как говорилось в п. 13, в результате эксперимента часто интересующая нас функция $y=f(x)$ оказывается заданной в табличном виде (2), и тогда может возникнуть вопрос о подборе для нее приближенной *эмпирической формулы.* При этом обычно начинают с того, что изображают значения функции на миллиметровке или иной приспособленной для этого бумаге. После этого выбирают вид формулы, которой будут пользоваться. Если этот вид не вытекает из каких-либо общих соображений, то обычно выбирают одну из функций, описанных в пп. 22—29, или простую комбинацию таких функций (сумму степенных или показательных функций и т. п.); конечно, для этого надо хорошо представлять себе графики этих функций. При этом следят за тем, чтобы подбираемая функция $\varphi(x)$ имела те же характерные особенности, что изучаемая функция $f(x)$: скажем, если по своему физическому смыслу $f(x)$ четная и $f(0)=0$, то и функция $\varphi(x)$ должна обладать этими свойствами и т. п. Иногда не удается подобрать единую формулу на всем интервале изменения x и приходится разбивать этот интервал на части и на каждой подбирать свою формулу.

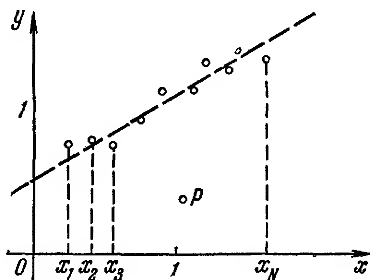


Рис. 49.

После выбора вида формулы нужно определить значения параметров, входящих в эту формулу.

Пусть, например, после построения точек получилась картина, изображенная на рис. 49. Если при эксперименте или при вычислениях не были исключены существенные ошибки, то точки, значительно выпадающие из общего хода зависимости, как точка P на чертеже, отбрасываются. Впрочем, иногда такие точки свидетельствуют о каких-то важных неучтенных факторах, и тогда их надо принять во внимание.

Оставшиеся точки на рис. 49 напоминают о линейной зависимости вида $y=ax+b$. Чтобы найти параметры a и b, проведем на чертеже прямую, к которой экспериментальные точки лежат ближе всего; это легко сделать, наложив на чертеж прозрачную линейку и передвинув ее на глаз в нужное положение. Так, на рис. 49 получаем $b=0{,}50$, $a=\frac{\Delta y}{\Delta x}=0{,}58$, т. е. $y=0{,}58x+0{,}50$.

Описанный подбор линейной зависимости сравнительно прост. Поэтому при выборе зависимости другого типа часто стараются так ввести новые переменные, чтобы в них зависимость стала линейной, после чего уже найти параметры, входящие в эту зависимость (это *метод выравнивания*). Конечно, так можно делать, если таких параметров не более двух, так как у линейной функции имеется два параметра.

Пусть, например, эксперимент привел к таблице значений:

x	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
y	0,00	0,01	0,03	0,08	0,17	0,29	0,45	0,66	0,91	1,22	1,57

Изображение экспериментальных точек на миллиметровке, которое мы предоставляем сделать читателю, напоминает о степенной функции вида $y=ax^\alpha$. Чтобы найти параметры a и α, прологарифмируем это равенство

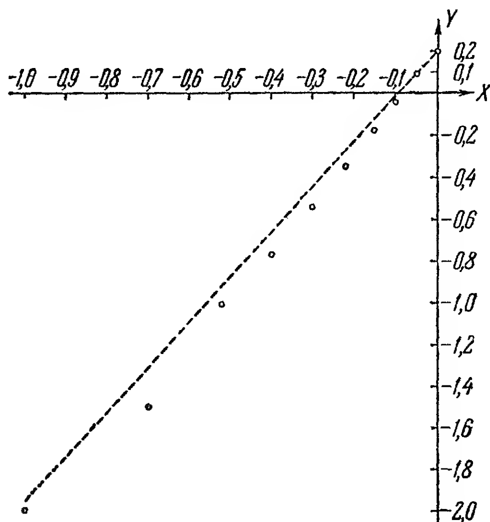


Рис. 50.

и обозначим $\lg y=Y$, $\lg x=X$, $\lg a=A$. Тогда мы приходим к равенству $Y=\alpha X+A$, т. е. в новых переменных зависимость является линейной. С помощью таблицы логарифмов построим таблицу значений новых переменных:

X	—1,0	—0,70	—0,52	—0,40	—0,30	—0,22	—0,15	—0,10	—0,05	0,00
Y	—2	—1,5	—1,1	—0,77	—0,54	—0,35	—0,18	—0,041	0,086	0,196

Полученные точки хорошо ложатся на прямую (рис. 50), при проведении которой надо больше ориентироваться на последние точки, известные с лучшей точностью. Из чертежа получаем $A=0{,}196$, $\alpha=2{,}44$, т. е. $a=1{,}57$ и окончательно получаем $y=1{,}57x^{2,44}$.

Дальнейшие указания и примеры см. в [21], § II.4.

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ НА ПЛОСКОСТИ

Аналитическая геометрия — это область математики, рассматривающая изучение геометрических задач средствами алгебры на основе метода координат.

§ 1. Координаты на плоскости

1. Декартовы координаты. Декартовы координаты известны из курса средней школы и уже применялись в гл. I. Они названы по имени Р. Декарта, который вместе с крупным французским математиком П. Ферма́ (1601—1665) является основателем метода координат.

Декартовы координаты нескольких точек показаны на рис. 51. Отметим, что оси координат здесь обязательно берутся взаимно перпендикулярными, единица масштаба — одинаковой по обеим осям, а начало отсчета по осям — в точке их взаимного пересечения (в гл. I при построении графиков функций это не всегда существенно, как указано в п. I.14). *Каждая точка на плоскости имеет вполне определенные координаты*, и, наоборот, *каждому набору координат x, y отвечает определенная точка на плоскости.* Это основное свойство дает возможность взамен точек рассматривать их координаты.

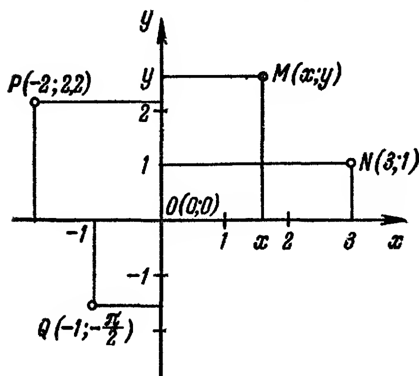


Рис. 51.

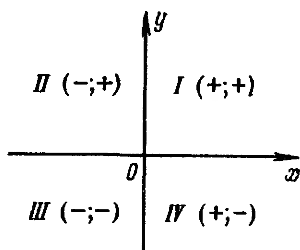


Рис. 52.

Оси координат делят плоскость на *четверти* (*квадра́нты*), которые нумеруются, как показано на рис. 52. Для каждой из этих

четвертей характерна своя комбинация знаков абсциссы и ординаты, также показанная на рис. 52.

2. Простые задачи на декартовы координаты. 1. *Расстояние между двумя точками.* Пусть даны точки $M_1(x_1;\ y_1)$ и $M_2(x_2;\ y_2)$ (т. е. даны координаты этих точек); требуется найти расстояние $d = M_1M_2$ (рис. 53). Формула для этого расстояния вытекает из теоремы Пифагора, примененной к прямоугольному треугольнику M_1M_2P. Имеем $M_1M_2^2 = M_1P^2 + PM_2^2$, т. е. $d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$ или

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \qquad (1)$$

Эта формула, как и дальнейшие, справедлива при любом расположении точек M_1 и M_2.

2. *Деление отрезка в данном отношении.* Пусть даны точки $M_1(x_1;\ y_1)$ и $M_2(x_2;\ y_2)$; требуется найти точку $M(x;\ y)$, лежащую

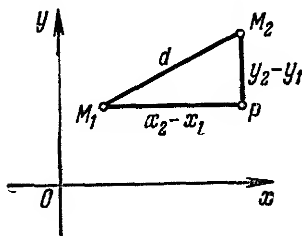


Рис. 53.

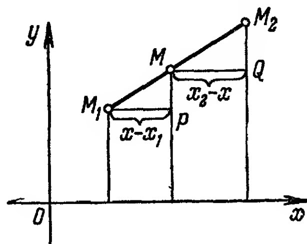


Рис 54.

на отрезке M_1M_2 и делящую его в данном отношении $\frac{M_1M}{MM_2} = \lambda$ (рис. 54). Решение этой задачи вытекает из подобия треугольников M_1PM и MQM_2, из которого следует, что $\frac{M_1P}{MQ} = \frac{M_1M}{MM_2} = \lambda$, т. е. $\frac{x - x_1}{x_2 - x} = \lambda$, $x - x_1 = \lambda x_2 - \lambda x$, откуда окончательно

$$x = \frac{x_1 + \lambda x_2}{1 + \lambda}; \quad y = \frac{y_1 + \lambda y_2}{1 + \lambda} \qquad (2)$$

(выражение для y выводится аналогично). В частности, если $\lambda = 1$, т. е. при делении отрезка M_1M_2 пополам, получается

$$x = \frac{x_1 + x_2}{2}; \quad y = \frac{y_1 + y_2}{2}.$$

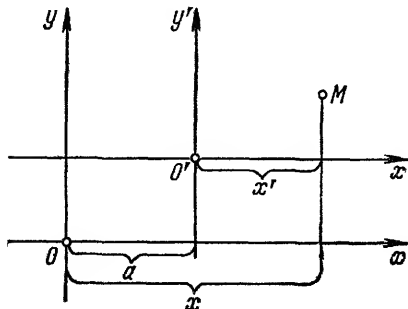


Рис. 55.

3. *Переход от одной декартовой системы координат к другой без изменения единицы масштаба.* Пусть на плоскости, помимо «старой» системы координат x, y, дана «новая» система координат x',

y'; требуется установить связь между старыми координатами и новыми. Мы рассмотрим три случая.

I. Пусть новые оси координат получаются в результате параллельного переноса старых, причем новое начало координат имеет старые координаты $(a;\ b)$. Тогда из рис. 55 получаем

$$x = x' + a; \quad y = y' + b.$$

II. Пусть новые оси получаются в результате зеркального отражения старых, например, относительно оси y. Тогда (рис. 56)

$$x = -x'; \quad y = y'. \tag{3}$$

III. Пусть новые оси получаются в результате поворота старых вокруг начала координат на угол α (рис. 57). Тогда из равенств $OC = OD - AB$, $CM = DB + AM$ получаем

$$\left.\begin{aligned} x &= x'\cos\alpha - y'\sin\alpha; \\ y &= x'\sin\alpha + y'\cos\alpha. \end{aligned}\right\} \tag{4}$$

Общий случай перехода от одной декартовой системы координат к другой сводится к комбинации разобранных частных случаев.

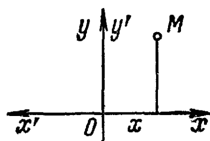


Рис. 56.

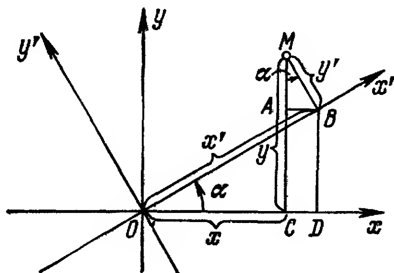


Рис. 57.

3. Полярные координаты. Помимо декартовых, на плоскости можно построить большое число других систем координат, т. е. способов охарактеризовать положение точки на плоскости с помощью двух числовых параметров (координат). Каждая из этих систем употребляется там, где это удобней, а декартова — чаще всех. Здесь мы рассмотрим только систему *полярных координат*, которая применяется, в частности, при исследовании вращательных движений. Для определения полярных координат надо выбрать *полюс* O и *полярную ось* Op (рис. 58), после чего положение точки M характеризуется *полярным радиусом* ρ, т. е. расстоянием от O до M и *полярным углом* φ (он же называется

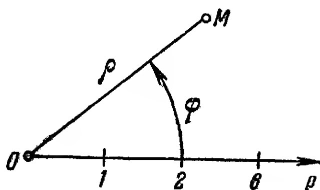


Рис. 58.

фазой точки M). При этом полярный угол считается положительным, если он откладывается в положительном направлении (как правило, против часовой стрелки). На рис. 59 показано построение нескольких точек по данным их полярным координатам; видно, в частности, что полюс имеет полярный радиус, равный нулю, и совершенно произвольный полярный угол. Для характеристики всех положений точки на плоскости достаточно только значений $-180° < \varphi \leqslant 180°$, однако иногда получаются значения φ за пределами этого интервала; при добавлении к полярному углу 360° положение точки не меняется.

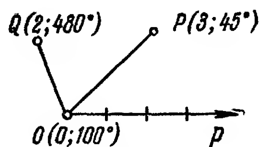


Рис. 59.

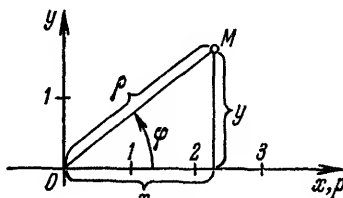


Рис. 60.

Связь декартовых координат с полярными, если они расположены, как на рис. 60, такова:

$$x = \rho\cos\varphi, \quad y = \rho\sin\varphi; \quad \text{наоборот,} \quad \rho = \sqrt{x^2+y^2}, \quad \operatorname{tg}\varphi = \frac{y}{x}. \quad (5)$$

§ 2. Линии на плоскости

4. Уравнение линии в декартовых координатах. Как мы видели в п. I.20, *уравнение*

$$F(x, y) = 0 \quad (6)$$

определяет на плоскости x, y (т. е. на плоскости, в которой задана декартова система координат x, y) *некоторую линию* (L), *представляющую собой совокупность всех точек, координаты которых удовлетворяют данному уравнению* (5); при этом соотношение (5) называется *уравнением линии* (L). Если, наоборот, первоначально дана линия (L) на плоскости x, y, то, формулируя аналитически геометрическое свойство, определяющее эту линию, мы получаем уравнение линии (L) в форме (5). (При этом надо иметь в виду, что каждое уравнение можно переписать в различных равносильных формах.) Это дает возможность взамен линий рассматривать их уравнения и тем самым сводить геометрические задачи к алгебраическим, которые, как правило, решаются значительно проще и единообразней, чем первые. Например, *чтобы проверить, проходит ли линия с уравнением* (6) (говорят просто «линия $F(x, y) = 0$») *через некоторую точку* (a; b), *достаточно подставить координаты этой точки в урав-*

нение линии и проверить, удовлетворится ли оно, т. е. будет ли $F(a, b)=0$.

Рассмотрим, например, *вывод уравнения окружности* (рис. 61). Пусть ее центр A имеет координаты $(a; b)$, а $M(x; y)$ — любая (текущая) точка на окружности. Тогда свойство, определяющее окружность, можно записать так: $AM=R$, где R — радиус данной окружности. Применяя формулу (1) для расстояния между двумя точками, получим $\sqrt{(x-a)^2+(y-b)^2}=R$ или, возведя в квадрат,

$$(x-a)^2+(y-b)^2-R^2=0.$$

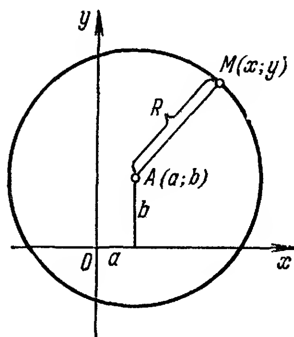


Рис. 61.

Это и есть соотношение, которому удовлетворяют координаты всех точек данной окружности и только этих точек, т. е. уравнение данной окружности. В нем a, b и R — фиксированные числа (параметры, определяющие положение и размеры окружности), а x и y — переменные *текущие координаты* точки окружности.

Пусть, наоборот, задано уравнение, например, вида

$$x^2+y^2-3x+4y-1=0. \quad (7)$$

При помощи дополнения до полного квадрата получаем

$$\left(x-\frac{3}{2}\right)^2-\left(\frac{3}{2}\right)^2+(y+2)^2-2^2-1=0, \qquad \left(x-\frac{3}{2}\right)^2+(y+2)^2-\frac{29}{4}=0.$$

Значит, данное уравнение — это уравнение окружности с центром в точке $(1{,}5; -2)$ и радиусом $\sqrt{\frac{29}{4}}=2{,}69$.

Если даны две линии с уравнениями $F_1(x, y)=0$ и $F_2(x, y)=0$, то может возникнуть задача о *нахождении точки пересечения* этих линий. Так как искомая точка должна принадлежать обеим линиям, то ее координаты x, y должны удовлетворять уравнениям обеих линий. Таким образом, для нахождения этих координат *надо решить систему уравнений*:

$$\left.\begin{aligned} F_1(x, y)&=0;\\ F_2(x, y)&=0. \end{aligned}\right\} \quad (8)$$

Число решений и даст число искомых точек пересечения; конечно, каждое решение состоит в указании значения x и значения y.

Пусть, например, требуется найти точку пересечения окружности (7) с прямой $y=x+b$, где b — некоторая постоянная. Для этого надо решить систему уравнений:

$$\left.\begin{aligned} x^2+y^2-3x+4y-1&=0;\\ y&=x+b. \end{aligned}\right\}$$

Если подставить y из второго уравнения в первое, раскрыть скобки и решить полученное квадратное уравнение относительно x, то после преобразований получим

$$x_1=\frac{-1-2b+\sqrt{9-28b-4b^2}}{4};\quad y_1=\frac{-1+2b+\sqrt{9-28b-4b^2}}{4};$$

$$x_2=\frac{-1-2b-\sqrt{9-28b-4b^2}}{4};\quad y_2=\frac{-1+2b-\sqrt{9-28b-4b^2}}{4}.$$

Посмотрим, при каком значении b обе точки пересечения совпадают. Для этого подкоренное значение должно равняться нулю, откуда получаем $b_{1,2}=\frac{-7\pm\sqrt{58}}{2}$, т. е. $b_1=0{,}31$, $b_2=-7{,}31$. При этих значениях b прямая $y=x+b$ касается заданной окружности (рис. 62). При $b_2<b<b_1$ точек пересечения две: $(x_1;\ y_1)$ и $(x_2;\ y_2)$. При остальных b прямая не пересекает окружность (подкоренное выражение оказывается отрицательным).

И в других примерах совпадение двух точек пересечения, координаты которых найдены из системы (8), обычно свидетельствует о том, что в этой общей точке обе заданные линии *соприкасаются* друг с другом, т. е. имеют в этой точке одну и ту же касательную.

Рис 62.

5. Уравнение линии в полярных координатах. В любой системе координат на плоскости уравнение между этими координатами определяет некоторую линию (исключения будут указаны в п. 8). Рассмотрим, в частности, полярные координаты. Мы будем считать, что уравнение разрешено относительно ρ, т. е. имеет вид

$$\rho=f(\varphi). \tag{9}$$

Придавая φ числовые значения и находя соответствующие значения ρ, получим точки, которые образуют линию на плоскости — график функции (9) в полярных координатах.

Рассмотрим два примера. График линейной зависимости $\rho=a\varphi+b$ изображен на рис. 63; это — *спираль Архимеда*. Она получается при наложении равномерного вращения и равномерного движения вдоль по радиусу, так как если

$$\rho=vt+b,\ \varphi=\omega t,\ \text{ то }\ \rho=\frac{v}{\omega}\varphi+b.$$

Итак, график одной и той же функции (в данном случае линейной) в полярных координатах и в декартовых (см. п. I. 22) имеет совершенно различный вид.

График показательной функции $\rho=e^{k\varphi}$ в полярных координатах изображен на рис. 64. Это — *логарифмическая спираль*. Она бесконечно накручивается на полюс, никогда его не достигая.

Логарифмическая спираль обладает рядом интересных свойств. Например, если ее подвергнуть *подобному преобразованию*, т. е. всестороннему равномерному растяжению в m раз (*m — коэффициент подобия*), то получится линия с уравнением $\rho = me^{k\varphi}$. Но $\rho = me^{k\varphi} = e^{k\left(\varphi + \frac{\ln m}{k}\right)} = e^{k(\varphi+\alpha)}$, где $\alpha = \frac{\ln m}{k}$, т. е. получится такой же результат, как если исходную спираль повернуть вокруг полюса по часовой стрелке на угол α радиан, так как

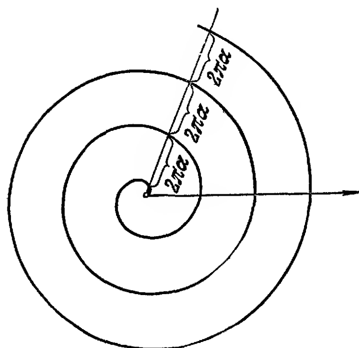


Рис. 63.

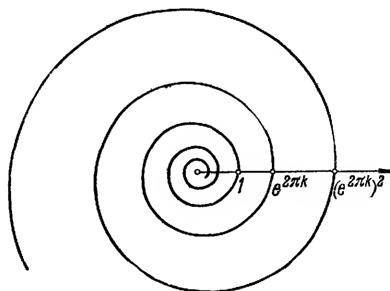
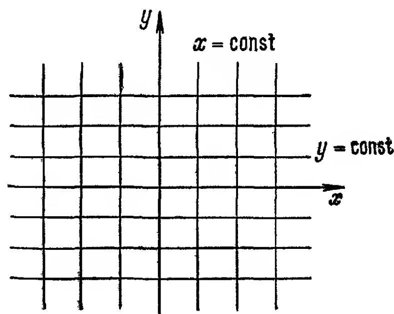


Рис. 64.

график $\rho = f(\varphi + \alpha)$ получается из графика $\rho = f(\varphi)$ поворотом вокруг полюса на угол α радиан в отрицательном направлении (почему?). Таким образом, логарифмическая спираль сама себе подобна с любым коэффициентом подобия Из других линий на плоскости этим свойством обладает только прямая.

В заключение скажем о *координатных линиях*, т. е. о линиях, на которых одна или другая координата сохраняют постоянное значение. В декартовых координатах координатные линии образуют два



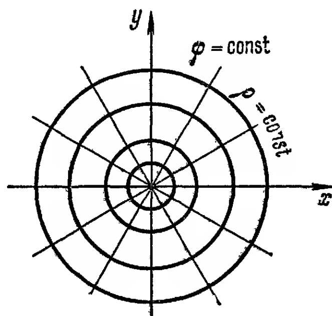


Рис. 65.

семейства прямых, параллельных одной или другой из осей координат. В полярных координатах линии $\rho = \text{const}$ образуют семейство концентрических окружностей, а линии $\varphi = \text{const}$ — семейство лучей, выходящих из полюса (рис. 65).

6. Параметрическое задание линий и функций. Бывает, что обе координаты, например декартовы, оказываются заданными как функции некоторой третьей переменной, которую мы обозначим буквой t:

$$x = \varphi(t); \quad y = \psi(t). \tag{10}$$

Эта переменная является параметром, определяющим положение точки $(x;\ y)$ на плоскости; когда t меняется, то точка на плоскости перемещается, описывая некоторую *линию* (L) (рис. 66), *заданную*, таким образом, *в параметрическом виде* (10). Так бывает, например, при исследовании движения точки в плоскости; тогда параметр t — время, формулы (10) определяют *закон движения*, а линия (L) называется *траекторией*. В других примерах параметр может иметь другой смысл, однако и

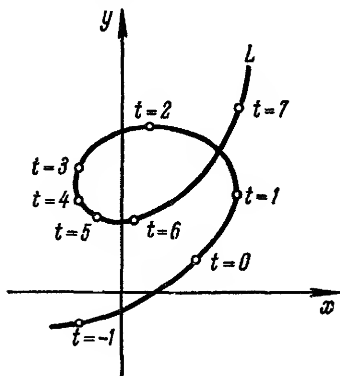


Рис. 66.

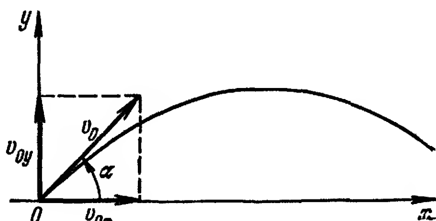


Рис. 67.

тогда его обычно можно для себя рассматривать так, как будто он является временем. Отметим, что одна и та же линия (L) может получиться при различных уравнениях (10), так как законы движения по одной и той же траектории могут быть различными (например, законы движения студентов вдоль траектории от трамвайной остановки до института за 22 и за 2 минуты до начала занятий).

Чтобы перейти к уравнению линии (L) в общей форме (6), надо из двух уравнений (10) исключить параметр (например, из первого уравнения выразить t через x и подставить результат во второе или как-нибудь иначе). Однако это не всегда возможно и не всегда целесообразно, так что часто приходится оставлять параметрическую форму.

Уравнения (10) задают определенную функциональную зависимость $y(x)$, так как если задать значение x, то из первого уравнения определится значение t (быть может, не одно), а из второго — соответствующее значение (или значения) y. Эта *функция* $y(x)$ оказывается *заданной в параметрическом виде*, а графиком ее служит линия (L).

Пример 1. Рассмотрим полет снаряда без учета сопротивления воздуха и шарообразности и вращения Земли (рис. 67). Если начальная скорость равна v_0, а угол бросания α, то горизонтальная составляющая скорости бу-

дет все время равна $v_{0x}=v_0\cos\alpha$, тогда как вертикальная составляющая скорости все время изменяется. Так как движение по вертикали происходит с постоянным ускорением g земного тяготения, то путь, пройденный по вертикали, отличается от свободного пути, соответствующего скорости $v_{0y}=v_0\sin\alpha$, на $\frac{gt^2}{2}$ (это доказывается в курсе механики). Таким образом,

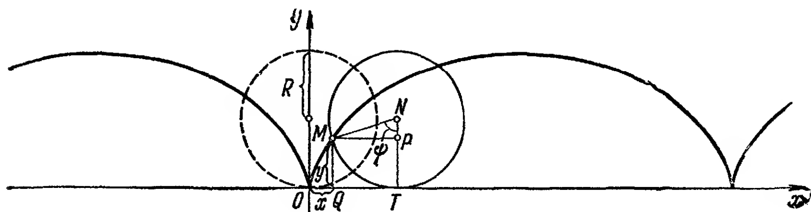


Рис 68.

получаем закон движения в плоскости x, y: $x=(v_0\cos\alpha)\,t$, $y=(v_0\sin\alpha)\,t-\frac{gt^2}{2}$. Это и есть закон движения, дающий одновременно траекторию в параметрическом виде. Исключая t, получим

$$y=(\operatorname{tg}\alpha)\,x-\frac{gx^2}{2v_0^2\cos^2\alpha}\,; \tag{11}$$

так как зависимость $y(x)$ квадратичная, то *траекторией будет служить парабола* (см. п. I. 23).

Пример 2. Рассмотрим траекторию точки окружности, катящейся без скольжения по прямой линии (на рис. 68 пунктиром показано положение

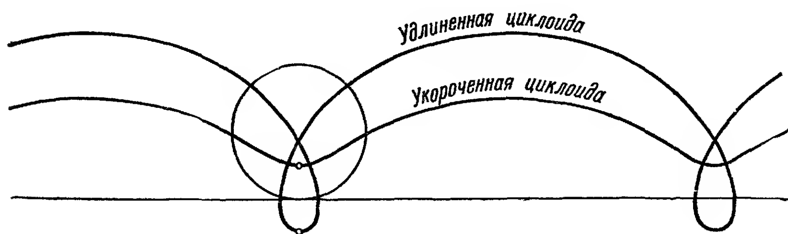


Рис. 69.

окружности в начальный момент, а сплошной линией — некоторое текущее положение). Выберем за параметр угол ψ поворота окружности. Тогда

$$\left.\begin{array}{l} x=OQ=OT-QT=\smile MT-MP=R\psi-R\sin\psi;\ x=R(\psi-\sin\psi);\\ y=QM=TN-PN=R-R\cos\psi;\quad y=R(1-\cos\psi); \end{array}\right\} \tag{12}$$

здесь отрезок OT приравнен дуге MT согласно условию отсутствия проскальзывания. Получились параметрические уравнения линии, которая называется

циклóидой, от греческого «киклос» — круг; она бесконечна и имеет точки возврата (рис. 68).

Циклоида является одной из простейших *рулéтт*, т. е. кривых, которые описывает на неподвижной плоскости точка, жестко связанная с одной из двух каких-либо линий, когда та без скольжения катится по другой. Другими примерами рулетт являются *укороченная* и *удлиненная циклоиды* (рис. 69), описываемые точкой, находящейся соответственно внутри или вне окружности, катящейся по прямой линии (при этом удлиненная циклоида имеет *точки самопересечения*); *гипоциклóиды* и *эпициклóиды*, которые описывает точка окружности, катящейся по другой окружности изнутри и извне (рис. 70), и т. д. Гипоциклоида при соотношении радиусов 1:4 называется *астрóидой*, от греческого «астрон» — звезда, а при 1:2 превращается в отрезок прямой; эпициклоида при 1:1 называется *кардиóидой* от греческого «кардиа» — сердце; эти линии и их уравнения, которые мы предоставим вывести читателю, показаны на рис. 71 и 72. Все эти линии имеют большое значение в теории механизмов.

Рис. 70.

7. Алгебраические линии. Если уравнение линии (L) в декартовых координатах x, y имеет вид

$$P(x,\ y)=0, \qquad (13)$$

где P — многочлен степени n, то говорят, что (L) — *алгебраическая линия порядка* n. Неалгебраическая линия называется трансцендентной. Так (см. § I. 22), график линейной функции, т. е. прямая линия, — это алгебраическая линия первого порядка, квадратичная па-

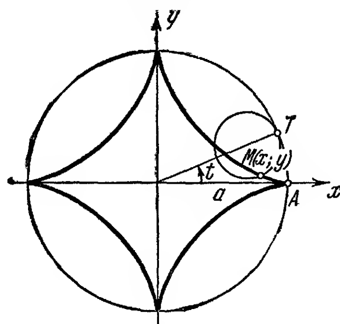


Рис. 71.

Астроида. Из равенства $\smile TM=\smile TA$ выведите уравнения $x=a\cos^3 t$, $y=a\sin^3 t$.

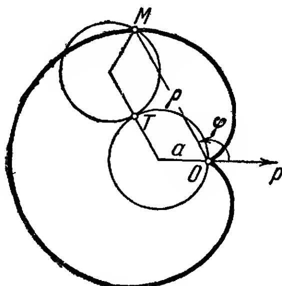


Рис. 72.

Кардиоида. Из равенства $\smile TM=\smile TO$ выведите уравнение $\rho=2a\,(1-\cos\varphi)$

рабола и окружность — линии второго порядка, кубическая парабола и полукубическая парабола (уравнение $y=x^{\frac{2}{3}}$ последней следует

переписать в виде $y^3-x^2=0$, чтобы получилось уравнение вида (13)) — это алгебраические линии третьего порядка. С другой стороны, синусоида, тангенсоида, график показательной функции — это трансцендентные линии.

В качестве интересного примера линий четвертого порядка рассмотрим *овалы Кассини*, введенные астрономом Кассйни (1625—1712). Такой овал определяется как совокупность точек плоскости, произведение расстояний которых от двух данных точек F_1 и F_2 этой плоскости есть величина постоянная. Расположив оси координат, как на рис. 73, и обозначив координаты точек $F_1(-a;\ 0)$, $F_2(a;\ 0)$, $M(x;\ y)$, а заданное произведение расстояний через b^2, получаем по формуле (1)

$$\sqrt{(x+a)^2+y^2}\cdot\sqrt{(x-a)^2+y^2}=b^2,$$

откуда легко вывести окончательное уравнение:

$$(x^2+y^2)^2-2a^2(x^2-y^2)=b^4-a^4.$$

В важном частном случае, когда $b=a$, линия имеет вид восьмерки и называется *лемнискатой*; она была открыта в 1694 г. швейцарским математиком Я. Бéрнулли (1654—1705). Перейдя к полярным координатам по формулам (5), получим уравнение лемнискаты $\rho^2=2a^2\cos 2\varphi$. При $0<b<a$ овал Кассини состоит из двух кусков.

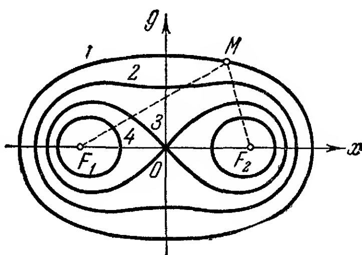


Рис. 73.

Надо отметить, что при определении порядка линии речь идет непременно об уравнении в декартовых координатах. Так, уравнение архимедовой спирали в полярных координатах (см. п. 5) имеет первую степень, но если его переписать в декартовых координатах $\sqrt{x^2+y^2}=a\operatorname{arctg}\frac{y}{x}+b$, то видно, что эта спираль—трансцендентная линия. Трансцендентны и все другие спирали и вообще все бесконечные линии, обладающие тем или иным свойством периодичности.

Порядок алгебраической линии не меняется при замене в плоскости одной декартовой системы координат на другую.

Действительно, например, при параллельном переносе (см. п. 2) уравнение данной линии приобретает вид

$$P(x'+a,\ y'+b)=0.$$

После раскрытия скобок и приведения подобных членов степень полученного многочлена не может получиться выше исходной (неоткуда). Могло бы показаться, что при приведении подобных членов общая степень многочлена может понизиться, если старшие члены взаимно уничтожатся. Но этого не может произойти, поскольку тогда при обратном переходе от x', y' к x, y степень многочлена должна была бы вновь повыситься, а мы только что видели, что этого не может быть. Аналогичная картина будет при преобразованиях другого вида.

Так как замена координат при неподвижной линии приводит к тому же результату, что и движение линии при неподвижных осях координат, то *порядок алгебраической линии остается инвариантным* (неизменным) *при движении этой линии как целого.*

Величина или вообще какой-либо объект, не меняющийся при тех или иных преобразованиях, называется *инвариантом* этих преобразований (или инвариантом относительно этих преобразований). Так, площадь является инвариантом движений, а углы — инвариантом не только движений, но также и преобразований подобия. Как мы видим, порядок алгебраической линии также является инвариантом движений. Важное понятие инварианта не знал некий владелец сада, который с невероятной быстротой красил забор, так как у него было очень мало краски и он хотел успеть покрасить забор, пока она не кончилась.

8. Особые случаи. Для некоторых уравнений вида $F(x, y)=0$ на плоскости x, y может получиться нечто, совсем непохожее на то, что принято называть линиями. Поясним это на примерах.

Уравнению $x^2+y^2+1=0$ на плоскости не удовлетворяет ни одна точка, так как при любых x, y левая часть положительная. Говорят, что это — *мнимая линия*, так как если допускать мнимые числа, то можно положить, например, $x=i$, $y=0$ и т. п. Но это название не меняет того, что данная линия, как подпоручик Киже в широко известном рассказе Ю. Тынянова, «фигуры не имеет».

Уравнению

$$x^2+y^2=0 \tag{14}$$

на плоскости удовлетворяет единственная точка: $x=0$, $y=0$ (начало координат). Если сравнить это уравнение с уравнением

$$x^2+y^2-R^2=0 \tag{15}$$

(см. п. 4), то получается как бы уравнение окружности нулевого радиуса. Вообще, если какой-либо объект зависит от параметров и при некоторых значениях этих параметров качественно меняется, теряя какие-либо весьма существенные свойства, то говорят, что при этих значениях происходит *вырождение* данного объекта. Окружность (15) зависит от параметра R; при $R=0$ она вырождается в точку (14), теряя важнейшее свойство — быть линией.

Уравнение алгебраической линии второго порядка

$$y^2-x^2=0 \tag{16}$$

можно переписать в виде

$$(y-x)(y+x)=0.$$

Но произведение равно нулю, если равен нулю какой-либо из множителей. Поэтому либо $y-x=0$, т. е. $y=x$, либо $y+x=0$, т. е. $y=-x$ (рис. 74). Каждое из этих уравнений определяет прямую линию на плоскости x, y. Таким образом, точка, удовлетворяющая уравнению (16), лежит либо на одной прямой, либо на другой. Итак, линия, определяемая уравнением (16), представляет

собой пару прямых, как говорят, *распадается* на пару прямых; линия второго порядка распалась на пару линий первого порядка.

Ни в коем случае нельзя считать, что гипербола (п. I. 25) тоже распадается; она состоит из двух кусков, каждый из которых представляет собой половину гиперболы, и оба они имеют единое уравнение. В случае же уравнения (16) каждая из полученных двух прямых имеет совершенно самостоятельное уравнение. Ясно, что таким путем можно объединить две совершенно произвольные линии, $F_1(x, y)=0$ и $F_2(x, y)=0$; для этого достаточно написать уравнение

$$F_1(x, y)\,F_2(x, y)=0.$$

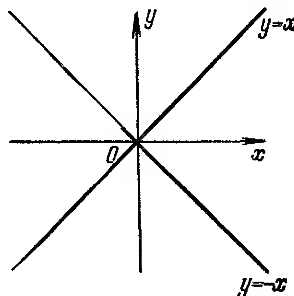


Рис. 74.

Также нельзя считать распадением разложение параболы рис. 23 на верхнюю и нижнюю половины на основе формулы

$$y^2-x=(y-\sqrt{x})(y+\sqrt{x}), \qquad (17)$$

которая вместо $y^2-x=0$ дает $y_{1,2}=\pm\sqrt{x}$. Дело в том, что распадение получается, когда многочлен в левой части (13) разлагается на произведение *многочленов*, а множители в правой части (17) не являются многочленами (см. п. I. 17).

§ 3. Алгебраические линии первых двух порядков

9. Линии первого порядка. Из п. 7 мы видим, что для получения линии первого порядка надо приравнять нулю многочлен первой степени. Он может содержать только члены первой степени и свободный член. Поэтому уравнение линии первого порядка в общем виде таково:

$$Ax+By+C=0. \qquad (18)$$

Здесь может быть два случая. Если $B\neq 0$, то, производя деление на B и обозначая

$$-\frac{A}{B}=k; \quad -\frac{C}{B}=b, \qquad (19)$$

получим

$$y=kx+b. \qquad (20)$$

В п. I. 22 мы видели, что это — уравнение прямой линии (там вместо k стояло a, что несущественно), изображенной на рис. 75. Если же $B=0$, то, деля на A и обозначая $-\frac{C}{A}=a$, получаем уравнение $x=a$, т. е. прямую, параллельную оси y. Отметим, что для таких прямых угловой коэффициент $k=\operatorname{tg}\frac{\pi}{2}=\pm\infty$, что также

вытекает из выражения (19), а уравнение записать в форме (20) невозможно.

Итак, *линии первого порядка — это прямые линии.*

Рассмотрим несколько простых задач на прямые линии.

1. *Через данную точку* $(x_1;\ y_1)$ *провести прямую с данным угловым коэффициентом* k. Конечно, в аналитической геометрии «провести прямую» означает «написать уравнение прямой». Искомое уравнение имеет вид (20), но b в нем неизвестно. Однако раз прямая проходит через данную точку, то координаты этой точки должны удовлетворять уравнению прямой: $y_1 = kx_1 + b$. Производя вычитание, исключаем b и получаем искомое уравнение

$$y - y_1 = k(x - x_1). \qquad (21)$$

Рис. 75.

Если в этом уравнении менять k, то мы получим пучок всевозможных прямых, проходящих через точку $(x_1;\ y_1)$. Можно положить и $k = \pm\infty$, т. е. получить вертикальную прямую; однако для этого надо предварительно обе части разделить на k, тогда после подстановки получится просто $0 = x - x_1$, т. е. $x = x_1$. Аналогичные предосторожности принимаются и в других задачах, когда параметры принимают бесконечные значения.

2. *Провести прямую через две данные точки* $(x_1;\ y_1)$ и $(x_2;\ y_2)$. Уравнение искомой прямой имеет вид (21), но k неизвестно. Однако из условия прохождения через вторую точку получаем $y_2 - y_1 = k(x_2 - x_1)$, откуда, производя деление, исключаем k:

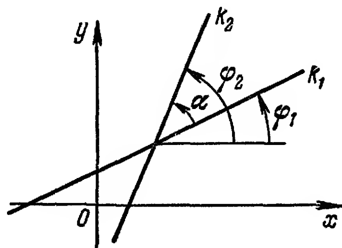


Рис. 76.

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}. \qquad (22)$$

Отметим, что в этом уравнении, как и в (21), x и y — это переменные координаты текущей (любой) точки искомой прямой.

3. *Найти угол между прямыми с данными угловыми коэффициентами* k_1 *и* k_2. Решение видно из рис. 76:

$$\operatorname{tg}\alpha = \operatorname{tg}(\varphi_2 - \varphi_1) = \frac{\operatorname{tg}\varphi_2 - \operatorname{tg}\varphi_1}{1 + \operatorname{tg}\varphi_1 \operatorname{tg}\varphi_2} = \frac{k_2 - k_1}{1 + k_1 k_2}. \qquad (23)$$

4. *Условие параллельности двух прямых* очевидно: $k_1 = k_2$.

5. *Условие перпендикулярности двух прямых* вытекает из задачи 3: для перпендикулярных прямых $\alpha=\frac{\pi}{2}$; но $\operatorname{tg}\frac{\pi}{2}=\pm\infty$, откуда $1+k_1k_2=0$ или $k_2=-\frac{1}{k_1}$.

10. Эллипс. *Эллипсом называется совокупность всех точек плоскости, сумма расстояний которых от двух данных точек этой плоскости* (называемых *фокусами* этого эллипса) *есть величина постоянная.* Из этого определения вытекает способ построения эллипса

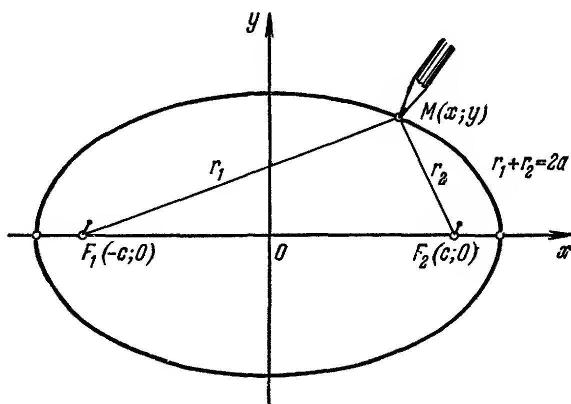


Рис. 77.

с помощью натянутой нити, показанный на рис. 77 и дающий представление о форме эллипса: это замкнутая выпуклая линия с двумя осями симметрии, называемыми *главными осями* эллипса, и с центром симметрии O, называемым *центром эллипса.*

Для вывода уравнения эллипса в наиболее простом виде расположим оси координат так, как показано на рис. 77, и обозначим $F_1F_2=2c$, $r_1+r_2\,(=\text{const})=2a$. Тогда с помощью формулы (1) для расстояния между точками уравнение эллипса можно написать в виде $\sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2}=2a$, откуда получаем последовательно

$$\sqrt{(x+c)^2+y^2}=2a-\sqrt{(x-c)^2+y^2};$$

$$(x+c)^2+y^2=4a^2-4a\sqrt{(x-c)^2+y^2}+(x-c)^2+y^2;$$

$$a\sqrt{(x-c)^2+y^2}=a^2-cx; \quad a^2\left[(x-c)^2+y^2\right]=a^4-2a^2cx+c^2x^2;$$

$$x^2(a^2-c^2)+a^2y^2=a^2(a^2-c^2). \tag{24}$$

Из треугольника F_1MF_2 видно, что $2a>2c$, т. е. $a^2-c^2>0$. Обозначим для краткости $a^2-c^2=b^2$. Тогда из равенства (24)

получим так называемое *каноническое уравнение эллипса*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{25}$$

Из этого уравнения вновь видно, что оси координат служат осями симметрии эллипса, так как если точка $(p;\ q)$ удовлетворяет уравнению (25), то и точки $(-p;\ q)$, $(-p;\ -q)$ и $(p;\ -q)$ тоже (рис. 78).

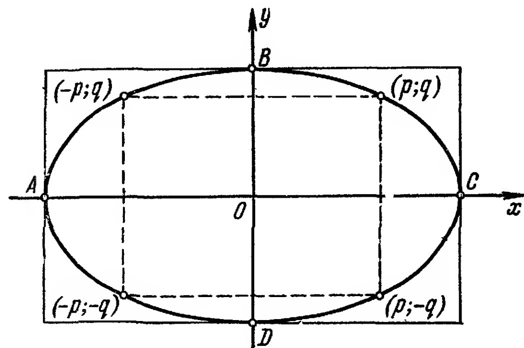


Рис. 78.

Если положить $y=0$, то получается $x=\pm a$, а если положить $x=0$, получается $y=\pm b$; значит, a и b — это длины *полуосей эллипса* (рис. 78) — *большой* $(AO=OC=a)$ и *малой* $(DO=OB=b)$. Кроме того, каждое из слагаемых в левой части (25) не может быть больше единицы, откуда $|x|\leqslant a$, $|y|\leqslant b$, и потому весь эллипс расположен внутри прямоугольника, изображенного на рис. 78. Точки A, B, C, D, в которых эллипс пересекается своими осями симметрии, называются *вершинами* эллипса. Эллипс имеет четыре вершины.

Отношение $\varepsilon=\frac{c}{a}=\frac{\sqrt{a^2-b^2}}{a}$, $0<\varepsilon<1$, называется *эксцентриситетом эллипса*. Это — величина безразмерная и при подобном преобразовании эллипса, когда все его размеры увеличиваются в k раз, не меняется, так как $\frac{kc}{ka}=\frac{c}{a}$.

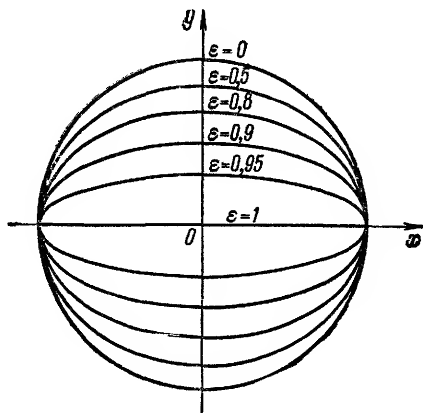


Рис. 79.

Эксцентриситет эллипса говорит о его форме (степени вытянутости), но ничего не говорит о его размерах. Интересно посмотреть, как влияет эксцентриситет на форму эллипса, если зафиксировать большую ось $2a$ и менять ε (рис. 79), имея в виду, что $c=\varepsilon a$, $b=a\sqrt{1-\varepsilon^2}$. При уменьшении ε фокусы сближаются, а малая

полуось приближается к большой. В пределе, когда $\varepsilon=0$, будет $c=0$, $b=a$, т. е. получится окружность. Таким образом, окружность является особым, предельным случаем эллипса, если фокусы слились в центр; тогда эксцентриситет равен нулю. Если, наоборот, ε приближается к 1, то эллипс становится все более вытянутым и в пределе вырождается в отрезок.

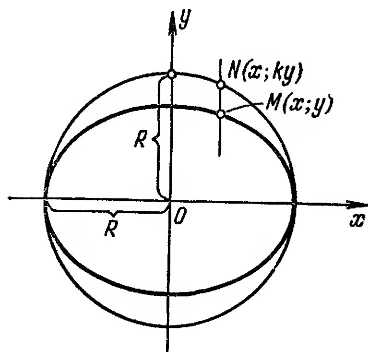


Рис. 80.

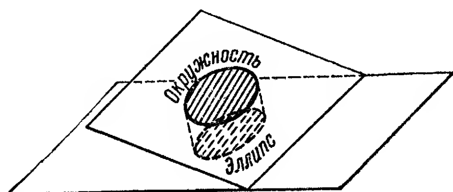


Рис. 81.

Эллипс *получается при равномерном сжатии окружности в одном направлении.* Действительно, рассмотрим, например, равномерное сжатие к оси x (рис. 80) в k раз. Если точка $M(x;\ y)$ лежит на кривой, полученной после сжатия, то точка $N(x;\ ky)$ должна лежать на окружности, откуда $x^2+(ky)^2=R^2$ или $\frac{x^2}{R^2}+\frac{y^2}{\left(\frac{R}{k}\right)^2}=1$, т. е. получается эллипс с полуосями R и $\frac{R}{k}$.

Из доказанного свойства легко вывести параметрические уравнения эллипса. Действительно, уравнения $x=R\cos t$, $y=R\sin t$ $(0\leqslant t\leqslant 2\pi)$ при заданном R определяют окружность радиуса R с центром в начале координат (почему?). Производя сжатие, получим $x=R\cos t$, $y=\frac{R\sin t}{k}$ или, вводя полуоси $a=R$, $b=\frac{R}{k}$, получим окончательно параметрические уравнения эллипса

$$x=a\cos t,\qquad y=b\sin t\qquad (0\leqslant t\leqslant 2\pi).\tag{26}$$

В п. XI.6 мы проверим, что при равномерном сжатии эллипса вновь получается эллипс.

Так как *при ортогональном проектировании какой-либо плоской фигуры она равномерно сжимается в одном направлении,* то при ортогональном проектировании окружности получается эллипс (рис. 81).

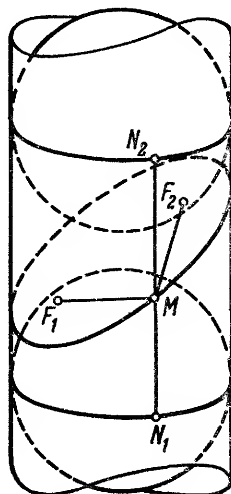


Рис. 82

Эллипс получается и при сечении прямого кругового цилиндра или конуса плоскостью. Для случая цилиндра это показано на рис. 82. Для доказательства того, что в сечении получится эллипс, в цилиндр вписываются шары так, чтобы они коснулись плоскости в точках F_1 и F_2. После этого для любой точки M сечения можно

написать, используя равенство касательных к шару, проведенных из одной и той же точки (рис. 82): $MF_1+MF_2=MN_1+MN_2=N_1N_2=\text{const}$; отсюда и следует наше утверждение. Для конуса построение аналогичное. Все эти свойства широко применяются в черчении.

11. Гипербола. Гипербола в нашем курсе уже встречалась в п. I.25. Однако мы на минуту об этом забудем, а потом, в п. 13, установим связь между пп. 11 и I.25. Поэтому дадим новое определение: *гиперболой называется совокупность всех точек плоскости, разность расстояний которых от двух данных точек этой плоскости* (называемых *фокусами* этой гиперболы) *есть величина постоянная.* Если выбрать оси координат, как на рис. 83, и обозначить $F_1F_2=2c$; $r_1-r_2=\pm 2a=$ $=\text{const}$, то получим уравнение гиперболы $\sqrt{(x+c)^2+y^2}-$ $-\sqrt{(x-c)^2+y^2}=\pm 2a$, откуда, проводя преобразования, как в п. 10, получим то же соотношение (24) (проверьте!). Однако теперь из треугольника F_1MF_2 видим, что $2a<2c$, и потому теперь нельзя, как в п. 10, обозначить $a^2-c^2=b^2$ (почему?), но можно обозначить $a^2-c^2=-b^2$. Тогда из (24) получаем

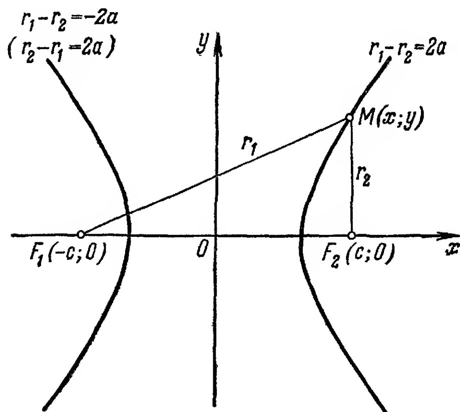


Рис. 83.

$$-b^2x^2+a^2y^2=-a^2b^2$$

и окончательно каноническое уравнение гиперболы:

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1. \tag{27}$$

Из этого уравнения видно, что и гипербола имеет две оси симметрии (*главные оси*), а также центр симметрии (*центр гиперболы*). Полагая $y=0$, получаем $x=\pm a$, а полагая $x=0$, получаем $y=\pm ib$.

Значит, ось x пересекает гиперболу в двух точках (*вершинах гиперболы*), это — *вещественная ось*; ось y гиперболу не пересекает, это — «*мнимая ось*». Соответственно постоянные a и b называются *вещественной* и *мнимой полуосями* гиперболы (хотя последнее название не совсем удачное, так как число b, конечно, вещественное).

Кроме того, из уравнения (27) видно, что $\frac{x^2}{a^2}\geqslant 1$, т. е. x либо $\leqslant -a$, либо $\geqslant a$ (рис. 84).

Гипербола имеет две асимптоты. Чтобы показать это, ограничимся первым квадрантом и запишем, используя (27):

$$y_{\text{гип}} = b\sqrt{\frac{x^2}{a^2} - 1} = \frac{b}{a}\sqrt{x^2 - a^2}.$$

Общие правила изучения подобных выражений при больших x будут указаны позже (см., в частности, конец п. IV.22). Пока мы не

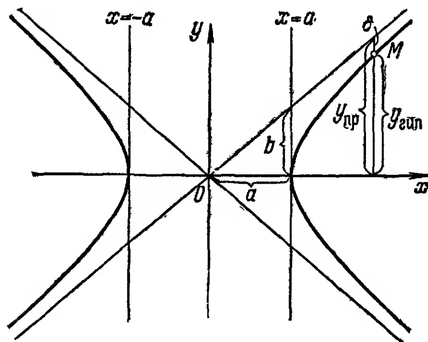


Рис. 84.

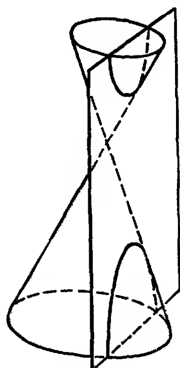


Рис. 85.

располагаем такими правилами, применим искусственное преобразование, «выделив» x из $\sqrt{x^2 - a^2}$,

$$\frac{b}{a}\sqrt{x^2 - a^2} = \frac{b}{a}\left[x + \left(\sqrt{x^2 - a^2} - x\right)\right] = \frac{b}{a}x + \frac{b}{a}\left(\sqrt{x^2 - a^2} - x\right).$$

Чтобы выяснить поведение слагаемого $\frac{b}{a}\left(\sqrt{x^2 - a^2} - x\right)$ при увеличении x, помножим и разделим это слагаемое на $\sqrt{x^2 - a^2} + x$; получим

$$y_{\text{гип}} = \frac{b}{a}x + \frac{b}{a}\,\frac{\left(\sqrt{x^2 - a^2} - x\right)\left(\sqrt{x^2 - a^2} + x\right)}{\sqrt{x^2 - a^2} + x} = \frac{b}{a}x - \frac{ab}{\sqrt{x^2 - a^2} + x}.$$

При удалении точки M по гиперболе в бесконечность последняя дробь, очевидно, безгранично приближается к нулю. Поэтому если рассмотреть прямую $y = \frac{b}{a}x$, то разность $\delta = y_{\text{пр}} - y_{\text{гип}}$ безгранично приближается к нулю и тем самым прямая является асимптотой гиперболы. Учитывая симметрию, получаем уравнение асимптот

$$y = \pm\frac{b}{a}x.$$

Можно проверить, что гипербола получается в результате пересечения бесконечного в обе стороны прямого кругового конуса плоскостью (рис. 85),

наклоненной к оси ближе, чем образующие конуса. (Попробуйте это доказать при помощи рассуждения, аналогичного проведенному в конце п. 10)

Определение гиперболы легко продемонстрировать в следующем примере. Пусть звуковой сигнал, поданный в точке A, уловили в паре пунктов B и C, причем в B на τ *сек* раньше, чем в C. Тогда относительно точки A можно гарантировать, что она находится на ближайшем к B куске гиперболы, имеющей фокусы в B и C и вещественную полуось $a=\frac{v_{зв}\tau}{2}$, где $v_{зв}$ — скорость звука (Почему?) Если произведено два таких замера в различных парах пунктов, то положение точки A определится пересечением соответствующих гипербол.

12. Родство эллипса, гиперболы и параболы. Между эллипсом (п. 10), гиперболой (п. 11) и параболой (п. I.23) имеется близкое родство. Это объясняется тем, что все они — линии второго порядка, причем, как мы увидим в п. 13, других линий второго порядка нет, если не принимать во внимание особых случаев, указанных в п. 8. Во многих задачах, включающих параметры, в решении может получиться одна из этих линий, в зависимости от значений параметров, причем парабола (или ее вырожденные состояния) всегда занимает промежуточное положение между эллипсом и гиперболой.

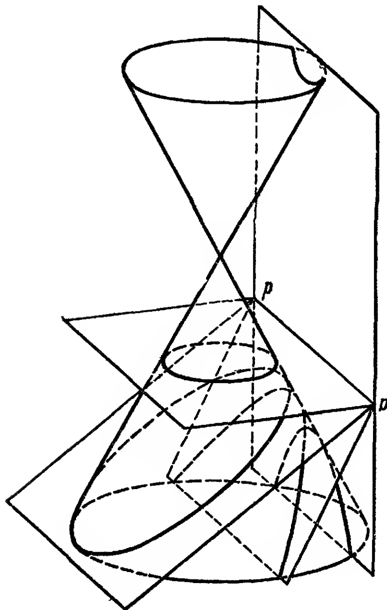


Рис. 86.

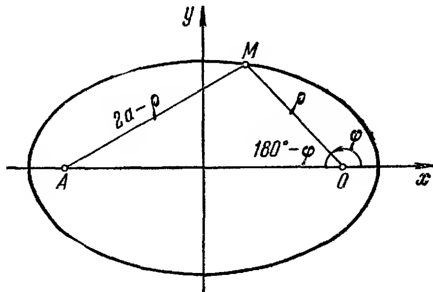


Рис. 87.

Рассмотрим *пересечение прямого кругового конуса с плоскостью*, поворачивающейся вокруг оси pp, выбранной, например, перпендикулярно к оси конуса (рис. 86). Пока наклон мал, в сечении получается эллипс. При увеличении наклона эллипс удлиняется, его эксцентриситет растет, а когда плоскость станет уже́ пересекать обе полы, в сечении получается гипербола. В промежуточном положении, когда плоскость наклонена к оси конуса так же, как образующие, линия пересечения бесконечна, но еще состоит из одного куска. Так как особых случаев, указанных в п. 8, здесь не будет, а в результате вырождения линии второго порядка не может получиться линия высшего порядка, то эта линия — парабола. По этой причине эллипс, гиперболу и параболу иногда называют *коническими сечениями*.

Рассмотрим с этой же точки зрения простейшее *полярное уравнение линий второго порядка*. Начнем с эллипса, для чего поместим полюс O в правый фокус (рис. 87). Применяя теорему косинусов к треугольнику AMO,

где A — левый фокус, а M — текущая точка эллипса, получим

$$AM^2 = AO^2 + OM^2 - 2AO\cdot OM\cos(180^\circ - \varphi);$$
$$(2a-\rho)^2 = (2c)^2 + \rho^2 + 2\cdot 2c\rho\cdot\cos\varphi;$$
$$4a^2 - 4a\rho + \rho^2 = 4(a^2 - b^2) + \rho^2 + 4c\rho\cos\varphi,$$

откуда

$$\rho = \frac{b^2}{a + c\cos\varphi} = \frac{\frac{b^2}{a}}{1 + \left(\frac{c}{a}\right)\cos\varphi}. \qquad (28)$$

Если для краткости обозначить $\frac{b^2}{a} = p$ (это — так называемый *«параметр»* эллипса), получим окончательно

$$\rho = \frac{p}{1 + \varepsilon\cos\varphi}. \qquad (29)$$

(Если поместить полюс в левый фокус, в знаменателе будет $1 - \varepsilon\cos\varphi$.) Если для гиперболы поместить полюс в левый фокус (рис. 88), то после аналогичных преобразований, которые мы предоставим читателю, получится та же формула (28), а если обозначить $\frac{b^2}{a} = p$, $\frac{c}{a} = \varepsilon$, то получится формула (29). Однако здесь безразмерная величина ε (также называемая *эксцентриситетом*) уже > 1.

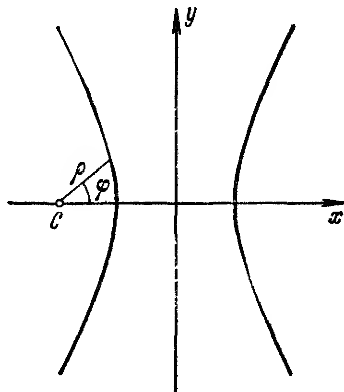


Рис. 88.

Легко проверить, переходя от полярных координат к декартовым по формулам (5), что при $\varepsilon = 1$ уравнение (29) представляет параболу; в самом деле, тогда можно написать $\rho = \frac{p}{1+\cos\varphi}$, $\rho + \rho\cos\varphi = p$, откуда $\sqrt{x^2+y^2} + x = p$, $\sqrt{x^2+y^2} = p - x$, $x^2 + y^2 = p^2 - 2px + x^2$ и окончательно

$$x = -\frac{1}{2p}y^2 + \frac{p}{2}$$

(см. п. 1.23). При этом полюс координат называется *фокусом* параболы; в отличие от эллипса и гиперболы, *парабола имеет только один фокус*. Интересно отметить, что на рис. 79 при $\varepsilon = 1$ из эллипса получился отрезок. Таким образом, в разных задачах вырождение может привести к различным результатам.

Уравнение (29) применяется, в частности, в небесной механике в *задаче двух тел*, движущихся под действием их взаимного притяжения. Рассмотрим, например, запуск искусственного спутника Земли из точки T, лежащей за пределами атмосферы, в горизонтальном направлении (рис. 89). Если начальная скорость v_0 недостаточна, то спутник вращаться вокруг Земли не будет. При достижении «первой космической скорости» спутник будет вращаться по круговой орбите, центр которой находится в центре Земли. Если v_0 увеличить, то оказывается, что вращение будет происходить по эллипсу, причем центр Земли будет находиться в одном из его фокусов.

При дальнейшем увеличении v_0 эксцентриситет эллипса будет возрастать, а его второй фокус удаляться При достижении «второй космической скорости» траектория станет параболической и спутник не вернется в T; таким

образом, парабола — это эллипс, у которого один из фокусов удален на бесконечность. При дальнейшем увеличении v_0 траектория станет гиперболической и второй фокус появится «с другой стороны»; центр Земли все время будет находиться в фокусе орбиты.

Конические сечения можно определить также следующим образом. Если уравнение (29) переписать в виде

$$\rho+\rho\varepsilon\cos\varphi=p, \quad \text{или} \quad \rho=p-\rho\varepsilon\cos\varphi=\varepsilon\left(\frac{p}{\varepsilon}-\rho\cos\varphi\right),$$

то полученное выражение в скобках как раз равно длине отрезка MM' на рис. 90 (почему?), где прямая ll проведена перпендикулярно к полярной

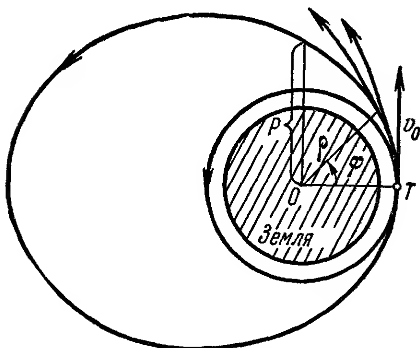


Рис. 89.

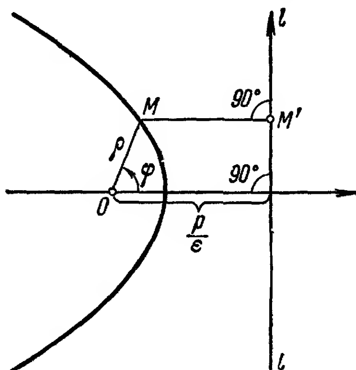


Рис. 90.

оси на расстоянии $\frac{p}{\varepsilon}$ от полюса. Но $\rho=OM$, т. е. получаем $OM=\varepsilon MM'$, откуда $\frac{MO}{MM'}=\varepsilon\ (=\text{const})$. Таким образом, эллипс, параболу и гиперболу можно определить еще как *совокупность всех точек, отношение расстояния которых от некоторой точки (фокуса) к расстоянию от некоторой прямой* (так называемой *директрисы*) *есть величина постоянная.*

13. Общее уравнение линии второго порядка. По аналогии с (18) напишем уравнение линии второго порядка в общем виде:

$$Ax^2+2Bxy+Cy^2+Dx+Ey+F=0 \tag{30}$$

($2B$, а не просто B, пишется только для большей простоты получающихся формул). Наша цель — при помощи замены декартовых координат (см. п. 2) суметь преобразовать уравнение (30) к более простому виду и тем самым разобраться, какую линию оно изображает.

Так как канонические уравнения не содержат членов с произведением координат, то постараемся прежде всего путем поворота осей устранить такой член. После поворота получим в силу формул (4) уравнение в новых координатах x', y':

$$A(x'\cos\alpha-y'\sin\alpha)^2+2B(x'\cos\alpha-y'\sin\alpha)(x'\sin\alpha+y'\cos\alpha)+\\+C(x'\sin\alpha+y'\cos\alpha)^2+D(x'\cos\alpha-y'\sin\alpha)+\\+E(x'\sin\alpha+y'\cos\alpha)+F=0. \tag{31}$$

Если здесь раскрыть скобки и собрать члены с произведением $x'y'$, то коэффициентом при $x'y'$ будет

$$-2A\cos\alpha\sin\alpha+2B\cos^2\alpha-2B\sin^2\alpha+\\+2C\sin\alpha\cos\alpha=2B\cos 2\alpha+(C-A)\sin 2\alpha.$$

Значит, должно быть

$$2B\cos 2\alpha+(C-A)\sin 2\alpha=0,\quad \text{т. е.}\quad \operatorname{tg} 2\alpha=\frac{2B}{A-C}. \qquad (32)$$

Отсюда и находим угол поворота осей α.

После поворота уравнение примет вид

$$A'x'^2+C'y'^2+D'x'+E'y'+F=0, \qquad (33)$$

где A', C', D', E' — новые коэффициенты, полученные из (31) после приведения подобных членов. Однако в п. XI.11 мы покажем, что при любом повороте осей, хотя коэффициенты A, B, C при членах второй степени, вообще говоря, меняются, но выражение $AC-B^2$ при этом остается инвариантным (неизменным). Так как в уравнении (33) нет члена с $x'y'$, т. е. $B'=0$, то отсюда получаем

$$A'C'\ (=A'C'-B'^2)=AC-B^2.$$

Таким образом, если выражение $AC-B^2$, составленное для исходного уравнения (30), положительно, то в уравнении (33) коэффициенты A' и C' имеют одинаковый знак, так как их произведение положительно; это — *эллиптический случай*. Если $AC-B^2<0$, то A' и C' имеют противоположные знаки (*гиперболический случай*). Если $AC-B^2=0$, то из коэффициентов A' и C' один равен нулю (*параболический случай*).

Остановимся на эллиптическом случае. Произведя в уравнении (33) дополнение до полного квадрата (ср. п. 4), получим уравнение вида

$$A'(x'-a)^2+C'(y'-b)^2+F'=0.$$

Производя параллельный перенос осей x', y' (см. п. 2), мы перейдем к новым координатам $x''=x'-a$, $y''=y'-b$. Уравнение примет вид

$$A'x''^2+C'y''^2=-F',\quad \text{т. е.}\quad \frac{x''^2}{-\frac{F'}{A'}}+\frac{y''^2}{-\frac{F'}{C'}}=1. \qquad (34)$$

Пусть для определенности A' и C' оба положительны. Тогда, если $F'<0$, мы получаем каноническое уравнение эллипса. Значит, и исходная линия была эллипсом, но относительно осей x, y смещенным и повернутым. Если же $F'>0$ или $F'=0$, то получаем особые случаи, указанные в п. 8: мнимую линию или точку.

Аналогично в гиперболическом случае линия будет гиперболой, а в параболическом — параболой, если не учитывать особых случаев, описанных в п. 8 и обычно не имеющих практического значения.

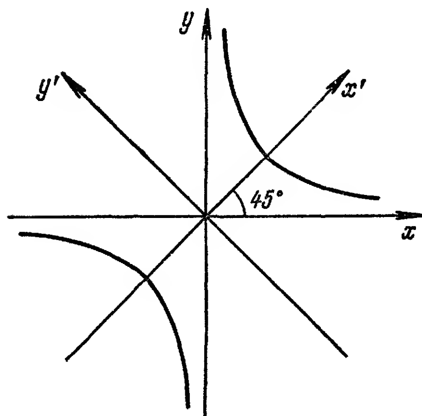


Рис. 91.

Рассмотрим, например, график обратной пропорциональности (см. п. I.25) с уравнением $y=\frac{k}{x}$, которое можно переписать в виде $xy-k=0$. Сравнивая с уравнением (30), видим, что здесь $A=C=D=E=0$; $B=\frac{1}{2}$; $F=-k$. Так как $AC-B^2=-\frac{1}{4}<0$, то это — гиперболический случай. По формуле (32) видим, что $\operatorname{tg} 2\alpha=\frac{1}{0}=\pm\infty$, откуда можно положить $2\alpha=\frac{\pi}{2}$, $\alpha=\frac{\pi}{4}$. Итак, повернем оси на 45°. По формулам (4) имеем

$$x=x'\frac{\sqrt{2}}{2}-y'\frac{\sqrt{2}}{2}=\frac{\sqrt{2}}{2}(x'-y'),$$
$$y=x'\frac{\sqrt{2}}{2}+y'\frac{\sqrt{2}}{2}=\frac{\sqrt{2}}{2}(x'+y').$$

Подставляя это выражение в исходное уравнение, получим

$$\frac{\sqrt{2}}{2}(x'-y')\frac{\sqrt{2}}{2}(x'+y')-k=0,$$

т. е.

$$\frac{x'^2-y'^2}{2}-k=0;\quad \frac{x'^2}{2k}-\frac{y'^2}{2k}=1;$$

значит, рассматриваемая линия — гипербола с равными полуосями ($a=b=\sqrt{2k}$) (рис. 91). Этим обосновывается употребление термина «гипербола» в п. I.25.

ПРЕДЕЛ. НЕПРЕРЫВНОСТЬ

§ 1. Бесконечно малые и бесконечно большие величины

1. Бесконечно малые величины. Бесконечно малые величины — это очень важный класс переменных величин, играющий первостепенную роль в высшей математике. *Переменная величина называется бесконечно малой в некотором процессе, если она в этом процессе безгранично приближается (стремится) к нулю.* Так, в процессе безграничного расширения данной массы газа плотность и давление будут величинами бесконечно малыми; это — пример положительной, непрерывной, монотонной, бесконечно малой величины (см. п. I.5). При затухающем колебании маятника угол его отклонения от положения равновесия в процессе течения времени также является бесконечно малой величиной, но эта величина уже колеблющаяся и вновь и вновь принимает значения обоих знаков, а также нулевое значение. При рассмотрении последовательности $a_1 = -\frac{1}{1^2}$, $a_2 = -\frac{1}{2^2}$, $a_3 = -\frac{1}{3^2}$, ... общий ее член $a_n = -\frac{1}{n^2}$ в процессе увеличения номера $n = 1, 2, 3, \ldots$ является бесконечно малой дискретной величиной, притом отрицательной, и т. д. Отметим, что при квалификации некоторой величины в качестве бесконечно малой непременно должен быть указан процесс, так как та же величина в другом процессе может уже вовсе не быть бесконечно малой.

Как мы сказали, бесконечно малая величина α должна «безгранично приближаться к нулю». Более подробно это расшифровывается так. В ходе развития процесса должен найтись момент, начиная с которого уже наверняка всегда будет $|\alpha| < 1$; некоторый другой, более поздний момент, начиная с которого всегда будет $|\alpha| < 0{,}1$; некоторый третий, еще более поздний момент, начиная с которого всегда будет $|\alpha| < 0{,}01$ и т. д. Это выражается такими словами: для *любого* заданного постоянного $\varepsilon > 0$ в ходе развития процесса должен найтись момент, начиная с которого всегда будет $|\alpha| < \varepsilon$. При этом нет надобности всегда такой момент фактически точно указывать: достаточно иметь уверенность, что он когда-либо наступит. Таким образом, бесконечно малая величина в начале своего изменения может быть вовсе не малой: существенно лишь, что она в ходе развития процесса становится *как угодно малой* (конечно, подразумевается, по абсолютной величине).

Уточним еще понятие «момент в развитии процесса». Если процесс рассматривается развивающимся во времени, то под «моментом» понимается просто момент времени. Однако развитие процесса может характеризоваться изменением не времени, а некоторой другой переменной величины (например, в третьем из приведенных примеров процесс состоит в том, что номер n принимает значения $1, 2, 3, \ldots$); тогда «момент» состоит в том, что эта величина принимает определенное значение.

Пользуясь уточнениями, содержащимися в последних двух абзацах, мы можем сказать, например, что общий член a_n какой-либо последовательности является бесконечно малым в процессе увеличения номера n, если для любого заданного $\varepsilon > 0$ должен найтись такой номер $N = N(\varepsilon)$, что при $n > N$ будет обязательно $|a_n| < \varepsilon$. Аналогично можно было бы уточнить понятие бесконечно малой и для других типов величин и процессов, однако это нам не понадобится.

С точки зрения приведенного выше определения постоянная величина, даже весьма малая, не является бесконечно малой; только постоянное число 0 с формальной точки зрения является бесконечно малой величиной.

Надо сказать, что это определение бесконечно малой величины, которым мы будем пользоваться, при его практическом применении приводит к следующему принципиальному затруднению: ни одна реальная величина не может *безгранично* приближаться к нулю. Действительно, в ранее приведенных примерах газ безгранично расширяться не может, а реальный маятник через некоторое время остановится. Рассматривая бесконечно малую массу, мы столкнемся с тем, что если взять массу уж очень малой, придется учитывать молекулярное строение вещества, а взять массу меньше массы элементарных частиц невозможно; то же и в других примерах.

Таким образом, указанное определение бесконечно малой можно применять лишь к *математической модели* реального процесса, в которой действительная картина изменена так, чтобы сделать это применение возможным: мы рассматриваем маятник, который может затухать бесконечно долго, материальные тела «сплошного» (немолекулярного) строения и т. п. Эта совершенно необходимая замена реального процесса на его математическую модель должна проводиться так, чтобы изучаемые стороны процесса не потерпели бы существенного искажения. Но модель есть лишь модель, и забвение этого может привести к принципиальным ошибкам, например, к попыткам навязать без должного обоснования все свойства моделей реальной действительности.

Есть и другой способ истолкования возможности практического применения бесконечно малых, и мы этим способом также будем пользоваться. Именно, *практическая бесконечность* различается от *математической бесконечности*. Так, «практическая» («физическая») бесконечно малая величина — это переменная или даже постоянная величина, достаточно малая по сравнению с участвующими «конечными» вели-

чинами (настолько малая, чтобы можно было без существенной ошибки применять по отношению к ней свойства «математических» бесконечно малых). В то же время эта величина не должна быть настолько малой, чтобы пришлось учитывать микроэффекты там, где это неуместно, или чтобы отрываться от реально возможных ее значений. Например, при изучении деформации упругого тела практически бесконечно малыми размерами следует считать размеры, достаточно малые по сравнению с размерами тела, но в то же время достаточно большие по сравнению с молекулярными размерами и т. п.

Ниже мы будем пользоваться определением, данным в начале этого пункта, однако время от времени будем вспоминать о высказанных здесь соображениях.

2. Свойства бесконечно малых. Свойства бесконечно малых сразу вытекают из их определения, данного в п. 1.

1. *Сумма или разность двух бесконечно малых есть также величина бесконечно малая.* Действительно, если каждое слагаемое безгранично приближается к нулю, то и сумма тоже. Аналогичным образом сумма трех, десяти и вообще любого ограниченного числа бесконечно малых есть также величина бесконечно малая. Отметим, что бывают случаи, когда в ходе рассматриваемого процесса число слагаемых в сумме безгранично растет; тогда, даже если каждое из слагаемых есть величина бесконечно малая, сумма может не быть бесконечно малой. Например,

$$\underbrace{\frac{1}{n^2}+\frac{1}{n^2}+\ldots+\frac{1}{n^2}}_{n\text{ раз}}=\frac{1}{n};\qquad \underbrace{\frac{3}{n}+\frac{3}{n}+\ldots+\frac{3}{n}}_{n\text{ раз}}=3;$$

$$\underbrace{\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{n}}+\ldots+\frac{1}{\sqrt{n}}}_{n\text{ раз}}=\sqrt{n}.$$

При увеличении n здесь будет как раз такая ситуация; в то же время первая сумма является величиной бесконечно малой, вторая — постоянной, а третья — даже безгранично возрастающей.

2. *Произведение бесконечно малой величины на величину ограниченную* (см. п. 1.5) *есть также величина бесконечно малая.* Пусть, например, первый множитель все время заключен в пределах от 0 до 1000, а второй последовательно принимает значения 1, 0,1, 0,01, 0,001 и т. д. Тогда значения произведения в эти моменты будут последовательно меньше $1000\times1=1000$, 100, 10, 1, 0,1, 0,01, 0,001 и т. д.

Из этого свойства вытекает, в частности, что *произведение бесконечно малой на величину постоянную есть величина бесконечно малая. Произведение двух бесконечно малых есть величина бесконечно малая*, так как бесконечно малая величина является, конечно, частным случаем ограниченной величины. Аналогичным образом произведение любого числа бесконечно малых есть величина бесконечно малая.

Отметим, что *частное от деления двух бесконечно малых может не быть бесконечно малым.* Если, например, $\alpha=\frac{1}{n}$, $\beta=\frac{1}{n^2}$, $\gamma=\frac{1}{n}+\frac{1}{n^2}$, где n принимает последовательные значения 1, 2, 3, ..., то величины α, β и γ — бесконечно малые. В то же время из их отношений $\frac{\beta}{\alpha}=\frac{1}{n}$; $\frac{\gamma}{\alpha}=1+\frac{1}{n}$; $\frac{\alpha}{\beta}=n$ первое является бесконечно малым, второе приближается к 1, а третье даже безгранично возрастает. О таких частных мы будем подробно говорить в § 3.

3. Бесконечно большие величины. *Переменная величина x называется бесконечно большой в некотором процессе, если она в этом процессе безгранично возрастает по абсолютной величине*; тогда пишут $|x|\to\infty$. Бесконечно большая величина может быть положительной ($x\to\infty$, иногда пишут $x\to+\infty$), отрицательной ($x\to-\infty$), но может также и менять знак: например, величина $x_n=(-2)^n$ при возрастании номера принимает значения -2, 4, -8, 16, ..., т. е. является бесконечно большой. Подробная расшифровка понятия «безгранично возрастает» аналогична той, которая была дана в п. 1 для понятия «безгранично приближается к нулю», но, конечно, надо рассматривать неравенства вида $|x|>N$. Это значит, что, начиная с некоторого момента, величина должна наверняка удовлетворять неравенству $|x|>1$, начиная с некоторого другого, более позднего момента, — неравенству $|x|>10$, начиная с третьего момента, — неравенству $|x|>100$ и т. д.

Отметим некоторые простые свойства бесконечно больших. *Величина, обратная бесконечно большой, является бесконечно малой, а величина, обратная бесконечно малой, является бесконечно большой.* Условно это записывают так:

$$\frac{1}{\infty}=0, \qquad \frac{1}{0}=\pm\infty.$$

Такие записи, которыми мы будем пользоваться, надо правильно понимать. Например, первая обозначает: если в равенстве $\frac{1}{x}=\alpha$ величина x безгранично возрастает, то в том же процессе величина α безгранично приближается к нулю (или, как в п. 1, если x — «практическая» бесконечно большая, то α — «практическая» бесконечно малая). Аналогично расшифровываются все формулы, содержащие знак бесконечности ∞. Так, формула $\operatorname{tg}\frac{\pi}{2}=\pm\infty$ является условной краткой записью следующего факта: в процессе, когда величина φ безгранично приближается к $\frac{\pi}{2}$, величина $x=\operatorname{tg}\varphi$ безгранично возрастает по абсолютной величине, т. е. является бесконечно большой, и т. п. Таким путем удается во многих случаях манипулировать со знаком ∞, *как* с обычным числом, хотя, конечно, ∞ не является конкретным

числом, но лишь значком для переменной бесконечно большой величины, причем в разных случаях разной.

Сумма бесконечно большой величины и величины ограниченной является величиной бесконечно большой, так как первое слагаемое «перетягивает». Сумма двух бесконечно больших одинакового знака есть также бесконечно большая. В отличие от этого сумма двух бесконечно больших противоположного знака может и не быть бесконечно большой, бесконечности могут «скомпенсироваться». Это записывают так: $\infty+\infty=\infty$; $\infty-\infty$ есть *неопределенность.* Последнее показывает, что со знаком ∞ не всегда можно действовать как с обычным числом: вовсе не всегда $\infty-\infty=0$, так как $\infty-\infty$ — это краткая запись разности $X-Y$, где X и Y — бесконечно большие, вообще говоря, различные, а тогда в разных конкретных примерах разность будет вести себя совершенно по-разному. Позже мы поговорим о неопределенностях более подробно.

Произведение двух бесконечно больших есть величина бесконечно большая. Более того, произведение бесконечно большой на величину, бóльшую по абсолютному значению некоторой положительной постоянной, есть величина бесконечно большая. В то же время частное от деления двух бесконечно больших, подобно частному от деления двух бесконечно малых, есть неопределенность.

§ 2. Пределы

4. Определение. Говорят, *что переменная величина x в некотором процессе стремится к конечному пределу a, если величина a постоянная и x в этом процессе безгранично приближается к a.* Тогда пишут

$$x\to a \quad \text{или} \quad \lim x=a$$

(lim — от латинского «limes», что значит «предел»).

Таким образом, конечным пределом *переменной* величины, если он имеется, служит величина *постоянная.*

Согласно данному определению *бесконечно малые величины — это величины, стремящиеся к нулю, т. е. имеющие пределом нуль. Бесконечно же большие величины конечного предела не имеют.*

Сказать «x безгранично приближается к a» — это все равно, что сказать «разность между x и a безгранично приближается к нулю», т. е. $x-a=\alpha$ есть бесконечно малая. Последнее же равенство можно переписать в виде

$$x=a+\alpha \quad \text{или} \quad x=(\lim x)+\text{б. м.}$$

Переменная величина x может стремиться к своему пределу a, оставаясь меньше его, т. е. со стороны меньших значений; тогда условно пишут $x\to a-0$ или $\lim x=a-0$ (это, конечно, условная

запись, так как $a-0=a$). Если x в процессе стремления к a остается больше a, то пишут $x \to a+0$. Наконец, x может стремиться к a, становясь вновь и вновь то больше, то меньше a, как при затухающих колебаниях. Все эти случаи показаны на рис. 92.

Можно подвести итог о видах переменных величин. Переменная величина x в некотором процессе может быть одного из следующих видов:

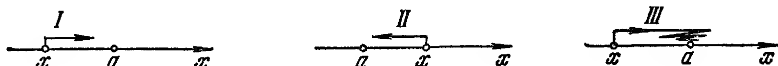


Рис. 92.

1) Ограниченная и притом имеющая предел; частным случаем, если предел равен нулю, является величина бесконечно малая. Для противопоставления иногда ограниченную величину называют *конечной* лишь в случае, если она не является бесконечно малой; например, можно говорить о бесконечно малой и конечной массах и т. п.

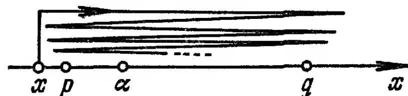


Рис. 93.

2) Ограниченная, но не имеющая предела; примером может служить отклонение маятника в случае незатухающих колебаний. Такая величина является колеблющейся (рис. 93).

В изображенном на рис. 93 случае любое значение α между p и q является *предельным значением* («*предельной точкой*») для величины x; это значит, что в развитии рассматриваемого процесса x в отдельные моменты безгранично приближается к α, но не обязано и далее удерживаться вблизи α. Среди этих предельных значений имеется наименьшее p и наибольшее q, которые обозначаются соответственно $\underline{\lim}\, x$ и $\overline{\lim}\, x$ и называются *нижним пределом* и *верхним пределом* величины x. Однако в рассматриваемом случае x *не имеет* единого предела, о котором говорилось в начале этого пункта. В связи с этим следует указать на то, что житейское понятие «предел» (в смысле «граница», «рубеж») отличается от приведенного выше математического понятия.

Ограниченная переменная величина x *всегда* имеет нижний и верхний пределы, причем $\underline{\lim}\, x \leqslant \overline{\lim}\, x$. Единый предел $\lim x$ имеется в том и только том случае, если $\underline{\lim}\, x = \overline{\lim}\, x$.

3) Неограниченная и притом бесконечно большая. Тогда пишут $\lim x = \pm\infty$ и говорят, что x имеет *бесконечный предел*.

4) Неограниченная, но не бесконечно большая; примером может служить отклонение колеблющегося тела от положения равновесия в случае резонанса. Такая величина также будет колеблющейся и то уходит «в бесконечность» все дальше и дальше, то возвращается на близкое расстояние от исходной точки (рис. 94).

5. Свойства пределов. 1. *Если рассматриваемая переменная величина имеет предел, то только один* (очевидно из определения).

2. *Предел постоянной величины равен ей самой* (очевидно).

3. Если в одном и том же процессе $x \to a$, а $y \to b$, то $x+y \to a+b$. Иначе это можно записать таким образом:

$$\lim(x+y) = \lim x + \lim y, \tag{1}$$

а сформулировать так: *предел суммы равен сумме пределов.* Для доказательства запишем $x = a + \alpha$; $y = b + \beta$, где α и β — бесконечно малые. Тогда $x + y = (a+b) + (\alpha+\beta)$. Значит, переменная величина $x+y$ представлена в виде постоянной $a+b$ и бесконечно малой $\alpha+\beta$ (см. п. 2). Следовательно, $(x+y) \to (a+b)$.

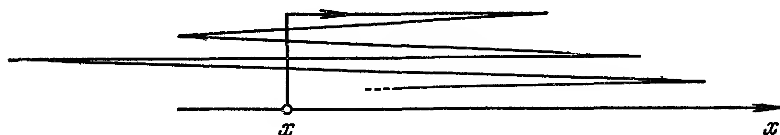


Рис. 94.

Полученный результат очень прост. Он говорит, по существу, что если $3{,}002 \approx 3$, а $2{,}001 \approx 2$, то $5{,}003 \approx 5$.

4. *Предел произведения равен произведению пределов*; более полная формулировка: *если сомножители имеют пределы, то и произведение имеет предел, который равен произведению пределов сомножителей.* Действительно, при старых обозначениях $xy = ab + (a\beta + b\alpha + \alpha\beta) \to ab$, т. е.

$$\lim(xy) = \lim x \lim y. \tag{2}$$

Здесь, как и в свойстве 3, мы взяли лишь две переменные величины, но легко проверить, что эти свойства сохраняются и для любого постоянного числа переменных величин. Например,

$$\lim(xyz) = \lim[(xy)z] = \lim(xy)\lim z = \lim x \lim y \lim z.$$

5. *Постоянный множитель можно выносить за знак предела*, т. е. $\lim(Cx) = C\lim x$, где $C = \text{const}$. Это вытекает из свойств 4 и 2.

6. *Предел частного равен частному от деления пределов*, т. е.

$$\lim\left(\frac{x}{y}\right) = \frac{\lim x}{\lim y}, \tag{3}$$

за исключением случаев, когда как числитель, так и знаменатель стремятся к нулю, т. е. в правой части получается неопределенное выражение $\frac{0}{0}$.

При доказательстве предположим сначала, что $\lim y = b \neq 0$. Тогда $\frac{x}{y} = \frac{a+\alpha}{b+\beta} = \frac{a}{b} + \left(\frac{a+\alpha}{b+\beta} - \frac{a}{b}\right) = \frac{a}{b} + \frac{\alpha b - \beta a}{b(b+\beta)}$. В последней дроби числитель бесконечно мал, а знаменатель $\approx b^2 = \text{const} \neq 0$, поэтому и вся вторая дробь бесконечно мала, а первая постоянна, откуда и вытекает наше утверждение.

Если же $\lim y = 0$, а $\lim x \neq 0$, то $\frac{x}{y} = \frac{1}{y} x \to \pm\infty$ (см. п. 3). Поэтому с обеих сторон формулы (3) получается $\pm\infty$.

В силу п. 3, формулы (1), (2) и (3) справедливы не только для конечных, но и для бесконечных пределов, за исключением случаев, когда в правых частях получаются неопределенные выражения вида $\infty - \infty$, $0 \cdot \infty$ и $\frac{\infty}{\infty}$, о которых будет разговор в § 3 и в § IV.4.

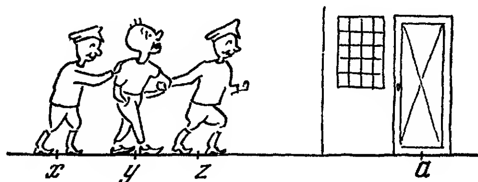


Рис. 95.

7. Если $x \to a$, причем $a > 0$, то $x > 0$, во всяком случае, если процесс зайдет достаточно далеко, т. е. начиная с некоторого момента. Это ясно из определения предела.

8. *В неравенстве можно перейти к пределу*: если $x \leqslant y$, то $\lim x \leqslant \lim y$ (естественно, если эти пределы существуют). Действительно, обозначим $z = y - x$. Тогда $z \geqslant 0$, поэтому z не может безгранично приближаться к постоянной отрицательной величине. Значит, $\lim z \geqslant 0$, $\lim (y - x) \geqslant 0$, $\lim y - \lim x \geqslant 0$.

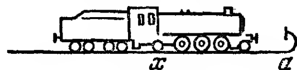


Рис. 96.

Заметим, что если $x < y$, то в пределе может получиться либо $\lim x < \lim y$, либо $\lim x = \lim y$, если разница между x и y в пределе сойдет на нет. Таким образом, пока не проведено дополнительное исследование, то строгое неравенство $x < y$ в пределе надо заменить на нестрогое $\lim x \leqslant \lim y$.

9. Если $x \leqslant y \leqslant z$ и в одном и том же процессе $x \to a$ и $z \to a$, то и $y \to a$. Это — так называемая «теорема о двух милиционерах». Продумайте это свойство, исходя из рис. 95.

10. Если величина x монотонно возрастает, то она либо возрастает безгранично, т. е. $x \to +\infty$, либо же она ограничена и тогда стремится к некоторому пределу, $x \to a - 0$. При этом, если x ограничено сверху постоянной величиной A, то и $\lim x = a \leqslant A$. Аналогично ведет себя монотонно убывающая величина. Это наглядное утверждение (рис. 96) на самом деле является очень глубоким свойством «полноты» совокупности всех вещественных чисел. Если бы мы пользовались только рациональными числами, то все предыдущие свойства пределов имели бы место, но не

свойство 10, так как из рациональных значений в пределе можно получить иррациональное. Строгое обоснование свойства 10 можно найти, например, в [48].

Таким образом, *ограниченная монотонная величина обязательно имеет конечный предел*; ограниченная немонотонная величина может предела не иметь (см. п. 4).

Заметим в заключение, что размерность величины при переходе к пределу сохраняется: если $x \to a$, то $[x]=[a]$.

Первую попытку создать теорию пределов сделал Ньютон в 1686 г., хотя операция предельного перехода фактически применялась и ранее, начиная с древнегреческих ученых. Близкое к приведенному здесь понятие предела сформулировал в 1765 г. французский математик и философ, просветитель периода подготовки буржуазной революции во Франции Ж. Даламбе́р (1717—1783).

6. Сумма числового ряда. Идея предельного перехода непосредственно применяется к важному понятию суммы ряда. Предварительно введем полезное сокращенное обозначение:

$$a_p+a_{p+1}+a_{p+2}+\ldots+a_{q-1}+a_q=\sum_{k=p}^{q} a_k. \qquad (4)$$

Здесь $\sum\limits_{k=p}^{q}$ ($\sum$ — греческая буква «сигма») — *знак суммирования*, который указывает, что в следующее за ним выражение надо подставить $k=p,\ p+1, \ldots, q$, а результаты сложить; a_k — *член суммы* (слагаемое), k — номер этого члена («индекс»), p и q — *нижний и верхний пределы суммирования*, указывающие границы изменения индекса. Например,

$$\sum_{k=3}^{8}\frac{1}{k^2}=\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}+\frac{1}{8^2}\,(=0{,}2774).$$

Отметим сразу, что *сумма не зависит от обозначения индекса суммирования*, т. е.

$$\sum_{k=p}^{q} a_k=\sum_{i=p}^{q} a_i=\sum_{j=p}^{q} a_j=\ldots$$

Действительно, все эти суммы равны левой части (4). Итак, *индекс суммирования является немым*, т. е. не входит в ответ и потому может быть обозначен любой буквой.

Перейдем теперь к «бесконечным суммам», точнее, к понятию числового ряда. *Числовым рядом* называется бесконечное выражение вида

$$a_1+a_2+\ldots+a_n+\ldots=\sum_{k=1}^{\infty} a_k; \qquad (5)$$

при этом слагаемые $a_1,\ a_2,\ a_3, \ldots$ — числа, называемые *членами ряда*. Чтобы определить понятие суммы ряда (5), надо сначала

составить *«частные суммы» ряда* (5):

$$S_1 = a_1; \quad S_2 = a_1 + a_2; \quad S_3 = a_1 + a_2 + a_3; \quad \ldots; \quad S_n = \sum_{k=1}^{n} a_k; \ldots$$

Если с возрастанием номера частные суммы стремятся к определенному конечному пределу, то ряд (5) называется *сходящимся*, а его *сумма* S полагается равной

$$S = \sum_{k=1}^{\infty} a_k = \lim_{n \to \infty} S_n = \lim_{n \to \infty} \sum_{k=1}^{n} a_k.$$

(Подпись маленькими буквами под знаком lim, равно как и под знаком $\to$, указывает на тот процесс, в котором берется рассматриваемый предел.) Для сходящегося ряда частные суммы с достаточно большими номерами практически просто равны друг другу и равны полной сумме ряда. Если конечного предела частных сумм нет, то ряд (5) называется *расходящимся*. В частности, если частные суммы стремятся к бесконечности, то ряд (5) называется *расходящимся к бесконечности*; тогда пишут

$$\sum_{k=1}^{\infty} a_k = \infty \qquad (\text{или } -\infty).$$

Расходящийся ряд конечной суммы не имеет.

Подобным образом определяется произведение бесконечного числа сомножителей и вообще результат любого бесконечного процесса: для этого сначала осуществляется конечный процесс, а затем производится предельный переход.

Рассмотрим, например, ряд

$$1 + \frac{1}{3} + \frac{1}{3^2} + \ldots + \frac{1}{3^n} + \ldots \tag{6}$$

Пользуясь формулой для суммы геометрической прогрессии, получим

$$S = \lim_{n \to \infty} S_n = \lim_{n \to \infty} \left(1 + \frac{1}{3} + \frac{1}{3^2} + \ldots + \frac{1}{3^{n-1}}\right) = \lim_{n \to \infty} \frac{1 - \frac{1}{3^n}}{1 - \frac{1}{3}} = \frac{1}{1 - \frac{1}{3}} = \frac{3}{2}.$$

Итак, ряд (6) сходится и его сумма равна 1,5. Например, частная сумма первых десяти членов приближенно равна 1,499975.

Подобным образом ряд

$$a + aq + aq^2 + \ldots + aq^n + \ldots \tag{7}$$

при $|q| < 1$ сходится и его сумма (сумма *бесконечно убывающей геометрической прогрессии*) равна $a(1-q)^{-1}$.

Для ряда

$$1 + 1 + 1 + \ldots + 1 + \ldots \tag{8}$$

частная сумма S_n равна n и потому стремится к бесконечности. Значит, ряд (8) расходится к бесконечности. Аналогично $-1-1-1-\ldots$ $\ldots-1-\ldots = -\infty$.

Для ряда

$$1-1+1-\ldots+(-1)^{n+1}+\ldots \tag{9}$$

частные суммы последовательно равны $S_1=1$, $S_2=0$, $S_3=1$, $S_4=0, \ldots$ и не имеют ни конечного, ни бесконечного пределов, а остаются ограниченными и колеблются без затухания (ср. конец п. 4, случай 2). Значит, ряд (9) расходится «колебательным» способом.

Отбрасывание или дописывание одного члена в ряде (5) не может нарушить факта сходимости или расходимости, т. е. если ряд (5) сходился, то он и будет сходиться, хотя сумма изменится, а если он расходился, то и будет расходиться. Действительно, если наряду с (5) рассмотреть ряд $a_2+a_3+\ldots+a_n+\ldots$, то его частная сумма отличается от соответствующей частной суммы ряда (5) на постоянное число a_1, и потому если одна из этих сумм стремится к пределу, то и другая тоже. Повторяя такое отбрасывание и дописывание членов, мы придем к выводу, что *произвольное изменение конечного числа членов ряда* (5) *не может нарушить факта сходимости или расходимости.*

Если ряд (5) сходится, то и ряд $R_n=a_{n+1}+a_{n+2}+a_{n+3}+\ldots$ сходится (почему?); его сумма называется *остатком* («хвостом») ряда (5). Ясно, что $S=\sum_{k=1}^{\infty} a_k=\sum_{k=1}^{n} a_k+\sum_{k=n+1}^{\infty} a_k=S_n+R_n$. Отсюда видно, что остаток сходящегося ряда, как разность между частной суммой и ее пределом, стремится к нулю с возрастанием номера.

Выведем *необходимый признак сходимости ряда* (5). Так как $S_{n-1}=a_1+a_2+\ldots+a_{n-1}$, $\quad S_n=a_1+a_2+\ldots+a_{n-1}+a_n$, то $a_n=S_n-S_{n-1}$, и потому *если ряд* (5) *сходится, то*

$$a_n \underset{n\to\infty}{\longrightarrow} S-S=0, \tag{10}$$

т. е. «общий член» a_n ряда (5) *с возрастанием номера стремится к нулю.* Этот признак не является достаточным для сходимости: например, для ряда $1+\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\ldots+\frac{1}{\sqrt{n}}+\ldots$ этот признак выполнен, но ряд расходится к бесконечности. (Эта расходимость следует из того, что $S_n=1+\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\ldots+\frac{1}{\sqrt{n}}>\frac{1}{\sqrt{n}}+$ $+\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{n}}+\ldots+\frac{1}{\sqrt{n}}=n\cdot\frac{1}{\sqrt{n}}=\sqrt{n} \underset{n\to\infty}{\longrightarrow} \infty$.)

Мы будем систематически изучать ряды различного вида в гл. XVII, где, в частности, будут указаны строгие признаки их сходимости. Однако отдельными рядами мы будем пользоваться раньше, поскольку для них вопрос о сходимости на практике обычно выясняют совсем просто, хотя и не совсем строго, следующим образом. Вычисляют члены один за другим, и *если они*

достаточно быстро выходят за пределы принятой точности вычисления, причем нет основания ожидать, что дальнейшие члены дадут в сумму существенный вклад, то все дальнейшие члены просто отбрасывают, т. е. заменяют ряд конечным числом его членов (частной суммой).

§ 3. Сравнение бесконечно малых

7. Сравнение бесконечно малых. Две бесконечно малые сравниваются друг с другом при помощи исследования их отношения. Пусть в некотором процессе рассматриваются бесконечно малые величины $\alpha \neq 0$ и $\beta \neq 0$. Тогда

1) если $\frac{\beta}{\alpha} \to 0$, то говорят, что β *быстрее стремится к нулю, чем* α, *или что* β *имеет высший порядок малости, чем* α, *а* α *имеет низший порядок малости, чем* β, и пишут $|\beta| \ll |\alpha|$ или $|\alpha| \gg |\beta|$; применяется также запись $\beta = o(\alpha)$. Таким образом, в данном случае β не просто бесконечно мала, но составляет бесконечно малую часть другой бесконечно малой α. Например, если ω — объем бесконечно малого кубика, а σ — объем столбика с таким же основанием, с постоянной высотой a, то $\omega \ll \sigma$, так как если обозначить через h длину ребра, то $h \to 0$ и $\omega/\sigma = h^3/ah^2 = h/a \to 0$;

2) если $\frac{\beta}{\alpha} \to \infty$, то $\frac{\alpha}{\beta} = \frac{1}{\beta/\alpha} \to 0$, т. е. $|\alpha| \ll |\beta|$;

3) если отношение $\frac{\beta}{\alpha}$ имеет конечный предел, отличный от нуля, то говорят, что α и β имеют *одинаковый порядок малости*; в этом случае ни одна из величин α, β не может стать чрезмерно меньше другой. В частности, если указанный предел равен единице, то величины α и β называются *эквивалентными*; тогда пишут $\alpha \sim \beta$. Например, если $x \to 0$, то величины x и $x + x^2$ эквивалентны; величины же $2x$ и $x + x^2$ имеют в этом процессе одинаковый порядок малости, но не эквивалентны, так как $2x/(x + x^2) \to 2$.

Можно проверить следующие свойства:

1) *если* α *и* β *имеют один порядок малости, а* $|\gamma| \ll |\alpha|$, *то* $|\gamma| \ll |\beta|$;

2) *если* α *и* β, *а также* β *и* γ *имеют один порядок малости, то и* α *и* γ *тоже*;

3) *если* α *и* β *имеют один порядок малости и одинаковый знак, то* $\alpha + \beta$ *имеет тот же порядок малости, что* α *и* β; если α и β имеют противоположные знаки, то $\alpha + \beta$ может получиться высшего порядка малости и т. п.

Проверим, например, первое свойство:

$$\lim \frac{\gamma}{\beta} = \lim \left(\frac{\gamma}{\alpha} \frac{\alpha}{\beta} \right) = \lim \frac{\gamma}{\alpha} \lim \frac{\alpha}{\beta} = 0 \cdot \lim \frac{\alpha}{\beta} = 0,$$

что и требуется. (Проверьте остальные свойства; впрочем, если вдуматься, то все они становятся очевидными без какого-либо доказательства.)

4) *Если* $\frac{\beta}{\alpha} \to k \neq 0$ *и обозначить* $\beta - k\alpha = \gamma$, *т. е.* $\beta = k\alpha + \gamma$, *то* $|\gamma| \ll |\alpha|$; другими словами, величины одного порядка малости прямо пропорциональны друг другу, с точностью до слагаемого высшего порядка малости. Это сразу следует из того, что

$$\frac{\gamma}{\alpha} = \frac{\beta - k\alpha}{\alpha} = \frac{\beta}{\alpha} - k \to 0.$$

Такой случай «почти пропорциональности» иногда обозначается так: $\alpha \backsim \beta$.

8. Свойства эквивалентных бесконечно малых. Справедливы следующие простые свойства:

1) *если* $\alpha \sim \beta$, *то* $\beta \sim \alpha$;

2) *если* $\alpha \sim \beta$ *и* $\beta \sim \gamma$, *то* $\alpha \sim \gamma$;

3) *если* $\alpha \sim \beta$, *то* $\alpha = \beta + \gamma$, *где* $|\gamma| \ll |\alpha|$ (и $|\gamma| \ll |\beta|$); другими словами, *эквивалентные бесконечно малые различаются на величину высшего порядка малости.* Наоборот, *если* $\alpha = \beta + \gamma$, *где* $|\gamma| \ll |\beta|$, *то* $\alpha \sim \beta$; другими словами, *прибавляя к бесконечно малой величине величину высшего порядка, получим величину, эквивалентную первой*;

4) *если* $\alpha \sim \alpha_1$ *и* $\beta \sim \beta_1$, *то* $\lim \frac{x\alpha}{y\beta} = \lim \frac{x\alpha_1}{y\beta_1}$, *где* x *и* y — *любые величины*; другими словами, *при нахождении предела можно бесконечно малые множители, стоящие в числителе и знаменателе, заменять эквивалентными величинами.*

Все эти свойства проверяются аналогично. Проверим, например, четвертое свойство:

$$\frac{x\alpha}{y\beta} = \frac{x\alpha_1}{y\beta_1} \frac{\alpha}{\alpha_1} \frac{\beta_1}{\beta}, \quad \text{откуда} \quad \lim \frac{x\alpha}{y\beta} = \lim \frac{x\alpha_1}{y\beta_1} \lim \frac{\alpha}{\alpha_1} \lim \frac{\beta_1}{\beta} = \lim \frac{x\alpha_1}{y\beta_1} \cdot 1 \cdot 1,$$

что и требуется.

9. Важные примеры. 1. *Длины бесконечно малой дуги и стягивающей ее хорды являются эквивалентными бесконечно малыми величинами,* т. е. $\overset{\frown}{MN}/MN \to 1$ при $N \to M$ (рис. 97). Это объясняется тем, что малая дуга почти не успевает изменить свое направление, т. е. «искривиться». В результате весьма малая хорда, увеличенная «под микроскопом» до постоянной конечной величины, практически неотличима от стягиваемой этой хордой дуги. При более строгом исследовании это наглядное свойство иногда принимается за аксиому, на которой основано само понятие длины дуги, а иногда выводится из аналогичных аксиом.

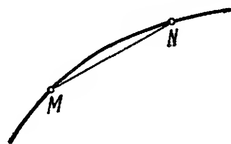


Рис. 97.

2. Применяя этот результат к дуге окружности (рис. 98), получим

$$\frac{MN}{\overset{\frown}{MN}}=\frac{2PN}{2\overset{\frown}{QN}}=\frac{2R\sin x}{2Rx}=\frac{\sin x}{x}\underset{x\to 0}{\longrightarrow}1.$$

Мы подразумевали, что $x>0$, но при перемене знака x выражение $(\sin x)/x$ не меняется, т. е. знак x не играет роли. Итак,

$$\lim_{x\to 0}\frac{\sin x}{x}=1. \qquad (11)$$

Заодно мы видим, что $\sin x<x$ при $x>0$ (так как $MN<\overset{\frown}{MN}$).

3. Воспроизведем рис. 41 с некоторыми дополнительными линиями (рис. 99). Мы видим, что $\operatorname{tg}\beta=\frac{\ln(1+h)}{h}$.

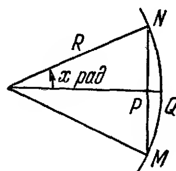


Рис. 98.

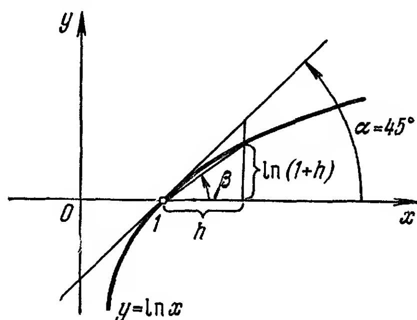


Рис 99.

Если же $h\to 0$, то $\beta\to\alpha=45°$, $\operatorname{tg}\beta\to\operatorname{tg}\alpha=1$, т. е.

$$\lim_{h\to 0}\frac{\ln(1+h)}{h}=1. \qquad (12)$$

На рис. 99 принято, что $h>0$, но тот же результат получится при $h<0$.

Приведем важное следствие из формулы (12). Так как $(1+h)^{\frac{1}{h}}=e^{\frac{\ln(1+h)}{h}}$, то при $h\to 0$ получаем

$$\frac{\ln(1+h)}{h}\to 1 \text{ и } e^{\frac{\ln(1+h)}{h}}\to e^1=e.$$

Итак,

$$\lim_{h\to 0}(1+h)^{\frac{1}{h}}=e. \qquad (13)$$

Последний предел *часто принимается за определение числа e.*

С помощью этих результатов можно вычислять многие другие пределы. Отметим, что в § IV.4 будет дан более стандартный и простой способ вычисления пределов.

10. Порядок малости. Если в одном процессе рассматриваются две бесконечно малые α и β, причем β имеет такой же порядок малости, что и α^k, то говорят, что β имеет *k-й порядок малости по сравнению с α*. Скорость стремления α к нулю служит как бы эталоном, с которым сравнивается скорость стремления β к нулю.

Примеры. Пусть $x \longrightarrow 0$, т. е. x—бесконечно малая величина, примем ее за эталон. Тогда, если $y=2x^2$, то y имеет второй порядок малости (по сравнению с x), так как y и x^2 имеют одинаковый порядок малости; если $z=4x^3+x^7$, то z имеет третий порядок малости, так как z и x^3 имеют одинаковый порядок малости ($z/x^3 \longrightarrow 4$). *Вообще порядок малости суммы (или разности) величин разного порядка определяется наименьшим из порядков малости слагаемых.* Именно член наименьшего порядка малости является *главным* в такой сумме; другими словами, все остальные члены бесконечно малы в сравнении с ним, а главный член почти полностью исчерпывает всю сумму, если процесс зайдет достаточно далеко. Далее, если $u=\sqrt{x}-x^2$, то u имеет половинный порядок малости и потому является величиной низшего порядка малости, чем x, т. е. $u/x \longrightarrow \infty$, $u \gg x$. Вообще, если порядок малости меньше единицы, то величина имеет низший порядок малости, чем эталон. Рассмотрим, наконец, $v=1-\cos x$; здесь v имеет второй порядок малости, так как

$$\lim_{x\to 0}\frac{v}{x^k}=\lim_{x\to 0}\frac{1-\cos x}{x^k}=\lim_{x\to 0}\frac{2\sin^2\frac{x}{2}}{x^k}=\lim_{x\to 0}\frac{2\sin\frac{x}{2}\sin\frac{x}{2}}{x^k}=\lim_{x\to 0}\frac{2\frac{x}{2}\frac{x}{2}}{x^k}$$

(в последнем переходе использовано четвертое свойство из п. 8), и поэтому, чтобы получился предел конечный и отличный от нуля, надо принять $k=2$.

При перемене эталона порядок малости может измениться, так что этот эталон необходимо указывать. Например, величина $y=x^6$ при $x \longrightarrow 0$ имеет шестой порядок по сравнению с x, но лишь второй порядок по сравнению с x^3.

11. Сравнение бесконечно больших. Сравнение бесконечно больших производится аналогично сравнению бесконечно малых. Некоторая разница имеется в терминах: так, если $\frac{x}{y} \longrightarrow 0$, где величины x и y—бесконечно большие, то говорят, что x имеет низший порядок по сравнению с y, а y—высший порядок по сравнению с x; но пишут все равно $|x| \ll |y|$ или $x=o(y)$. (Эти записи применяются во всех случаях, когда $x/y \longrightarrow 0$, даже если величины x и y и не являются бесконечно большими; отметим еще запись $x=O(y)$, которая означает, что отношение x/y ограничено.) С аналогичными незначительными изменениями переносятся все утверждения пп. 7, 8 и 10.

Отметим в заключение очевидное свойство: если $\lim x=0$; $\lim y=\text{const}\neq 0$; $\lim|z|=\infty$, то $|x|\ll|y|$, а $|y|\ll|z|$.

§ 4. Непрерывные и разрывные функции

12. Развернутое определение непрерывной функции. В п. I.16 уже было дано определение непрерывной функции. Сейчас мы более детально его расшифруем.

Пусть дана функция $y=f(x)$, и аргумент, принимавший сначала значение x_0, получил затем приращение Δx, т. е. $x=x_0+\Delta x$ (по поводу обозначений см. п. I.22). Тогда функция получит приращение

$$\Delta y = y - y_0 = f(x) - f(x_0) = f(x_0+\Delta x) - f(x_0). \qquad (14)$$

Функция f называется *непрерывной в точке* x_0 (т. е. при значении аргумента, равном x_0), *если в процессе, когда* $\Delta x \to 0$, *будет и* $\Delta y \to 0$, другими словами, *если бесконечно малому приращению аргумента отвечает бесконечно малое приращение функции*. В противном случае x_0 называется *точкой разрыва* функции f.

Так как (см. (14)) $f(x)=f(x_0)+\Delta y$, то условие $\Delta y \to 0$ равносильно такому: $f(x) \to f(x_0)$. Написав, далее, $x \to x_0$ вместо $\Delta x \to 0$, получим равносильные формулировки определения непрерывности:

$$f(x) \underset{x \to x_0}{\longrightarrow} f(x_0) \quad \text{или} \quad \lim_{x \to x_0} f(x) = f(x_0)$$

или, наконец,

$$\lim_{x \to x_0} f(x) = f(\lim x). \qquad (15)$$

Последнюю формулу читают так: *предел функции равен функции от предела.*

Подчеркнем, что значение функции в точке непрерывности *не может быть бесконечным.*

Функция, непрерывная в каждой точке некоторого интервала, называется непрерывной в этом интервале.

13. Точки разрыва. Если x_0 — точка разрыва функции f, то чаще всего само значение $f(x_0)$ бывает неопределенным, да обычно и не играет никакой роли. Важную роль играют пределы значений $f(x)$ при $x \to x_0-0$ и $x \to x_0+0$ (см. п. 4). Эти пределы обозначаются условно через $f(x_0-0)$ и $f(x_0+0)$ (рис. 100).

Бывает так, что указанные пределы конечны и $f(x_0-0)=f(x_0+0)$, но само значение $f(x_0)$ не определено или определено, но не совпадает с $f(x_0 \pm 0)$. Такой разрыв называется *устранимым*, так как если положить $f(x_0)=f(x_0 \pm 0)$ (это значение называется *«истинным» значением* функции $f(x)$ при $x=x_0$), то никакого разрыва больше не будет. Приведем простой пример: функция $f(x)$, заданная формулой

$$f(x) = \frac{\sin x}{x}, \qquad (16)$$

при всех $x \neq 0$ непрерывна, но при $x = 0$ не определена, так как подставить $x = 0$ в формулу (16) нельзя, получится неопределенное выражение 0/0. Однако если к формуле (16) добавить, что $f(0) = 1$, то в силу формулы (11) полученная функция $f(x)$ будет определена и непрерывна для всех x без исключения; при $x = 0$ был устранимый разрыв. Геометрически это можно представить так, что к линии pp (рис. 101) добавили одну точку M, после чего линия стала сплошной.

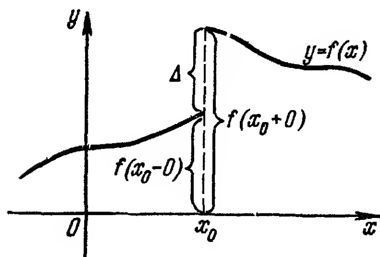


Рис. 100.

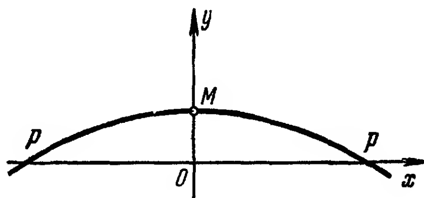


Рис. 101.

Если значения $f(x_0 - 0)$ и $f(x_0 + 0)$ конечны, но $f(x_0 - 0) \neq$ $\neq f(x_0 + 0)$, то говорят, что функция f имеет в точке x_0 *разрыв 1-го рода* или, что то же, *конечный скачок* $\Delta = f(x_0 + 0) - f(x_0 - 0)$ (см. рис. 100). Если же из двух значений $f(x_0 - 0)$ и $f(x_0 + 0)$ по

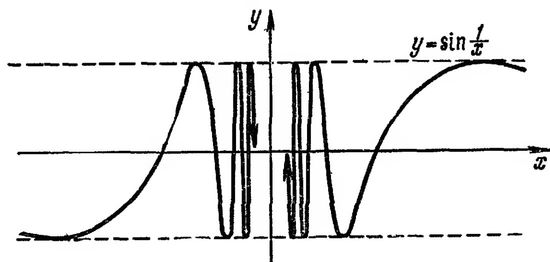


Рис. 102.

крайней мере одно обращается в бесконечность, то несколько неточно говорят, что функция f в точке x_0 *обращается в бесконечность* («уходит в бесконечность»); например, так ведет себя функция $f(x) = \frac{1}{x}$ при $x_0 = 0$.

Отметим в заключение, что бывает, хотя и очень редко, что $f(x_0 - 0)$ или $f(x_0 + 0)$ не имеет ни конечного, ни бесконечного значений, так как переменная величина может не иметь ни конечного, ни бесконечного пределов. Например, для функции $f(x) = \sin\frac{1}{x}$ при $x \to 0$ будет $\frac{1}{x} \to \infty$ и $\sin\frac{1}{x}$ будет вновь и вновь переходить от -1 к $+1$ и обратно, не имея предела (рис. 102).

Разрывы у физических переменных величин получаются при внезапном присоединении или отсоединении какого-либо воздействия, при переходе из одной среды в другую (на границе раздела), при внезапной перемене закона зависимости и т. п.

Так, на рис. 103 показано изменение тока в цепи, т. е. закон зависимости тока i от времени t, когда радист передает букву «а» по азбуке Морзе («точка — тире»). Как видим, получается функция, имеющая четыре точки разрыва, в каждой из которых она имеет конечный скачок, полученный за счет включения или отключения постоянной э. д. с. в цепи. Полезно обратить внимание на то, что более тщательный анализ (как говорят, применение «лупы времени», что сводится к значительному увеличению масштаба по оси t) показывает, что нарастание тока на самом деле происходит примерно так, как показано на рис. 104. Непременное наличие индуктивности в цепи приводит к тому, что разрыва функции $i(t)$ *на самом деле нет*, ток нарастает непрерывно, хотя

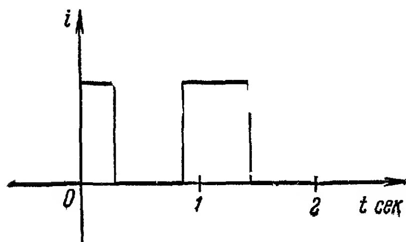


Рис. 103.

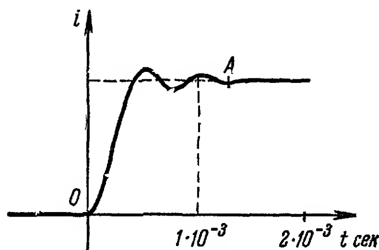


Рис. 104.

и очень быстро! И в некоторых случаях, например, если продолжительность импульса очень мала, учет непрерывности этого *переходного режима* (т. е., например, участка OA графика) очень важен. Однако если переходный режим значения не имеет, то проще схематизировать процесс, считая зависимость $i(t)$ разрывной в соответствии с рис. 103, если это не приводит к существенным ошибкам. Итак, одна и та же зависимость $i(t)$ является непрерывной или разрывной в зависимости от подхода (учета или неучета переходного режима)! При переходе из одной среды в другую аналогичную роль играет учет поверхностных эффектов и т. п.

Если рассматривается элементарная функция $f(x)$, то, как будет видно из п. 14, разрыв при $x=x_0$ может получиться только в том случае, если при подстановке $x=x_0$ где-нибудь в $f(x_0)$ получается выражение вида $\frac{a}{0}$, $\ln 0$ или 0^0. При этом оказывается, что предельные значения $f(x_0-0)$ и $f(x_0+0)$, конечные или бесконечные, существуют, если с соответствующей стороны от x_0 функция f определена; исключением может быть только случай, когда в выражении $f(x_0)$ присутствует $\sin\infty$ или $\cos\infty$.

Если функция f определена только с одной стороны от x_0, например справа, то может существовать только значение $f(x_0+0)$ («*концевое значение*»), а $f(x_0-0)$ нет. Роль концевых значений играют также пределы $f(-\infty)$ и $f(\infty)$.

14. Свойства непрерывных функций. 1. Сумма двух непрерывных функций является непрерывной функцией. Действительно, если функ-

ции $f_1(x)$ и $f_2(x)$ непрерывны, а $f(x)=f_1(x)+f_2(x)$, то при $x\to x_0$

$$\lim f(x)=\lim[f_1(x)+f_2(x)]=$$
$$=\lim f_1(x)+\lim f_2(x)=f_1(x_0)+f_2(x_0)=f(x_0),$$

что и доказывает непрерывность функции $f(x)$ (см. п. 12). Отметим, что при этом доказательстве мы сначала воспользовались свойством пределов (см. п. 5), а затем непрерывностью функций f_1 и f_2. Аналогичное применение других свойств пределов показывает, что:

сумма, разность и произведение любого конечного числа непрерывных функций есть снова непрерывная функция;

частное от деления двух непрерывных функций есть непрерывная функция всюду, где знаменатель отличен от нуля. В точках, где знаменатель обращается в нуль, частное или обращается в бесконечность, или становится неопределенным вида $\frac{0}{0}$, так что в обоих случаях непрерывность нарушается.

2. *Сложная функция, составленная из непрерывных функций, сама является непрерывной функцией.* Действительно, если функции $z(y)$ и $y(x)$ непрерывные и мы дадим x бесконечно малое приращение, то в силу непрерывности второй функции приращение y будет также бесконечно малым, а потому в силу непрерывности первой функции приращение z будет бесконечно малым; итак, сложная функция $z(x)$ непрерывная.

Из первых двух свойств можно сделать следующий вывод о непрерывности элементарных функций. Обзор простейших элементарных функций (см. п. I.18 и § I.4) показывает, что из них имеют разрыв только $y=x^{-m}$ для $-m<0$ при $x=0$ (когда получается выражение вида $1/0$), $y=\log_a x$ при $x=0$ (когда получается $\log 0$) и $y=\operatorname{tg} x$ при $x=\pm\pi/2,\ 3\pi/2,\ \ldots$ $\left(\text{когда получается } \sin\frac{\pi}{2}\Big/\cos\frac{\pi}{2}=\frac{1}{0}\right)$. Когда образуются сложные функции из простейших функций и их алгебраические комбинации, то в силу свойств 1 и 2 и п. 15.1 новые точки разрыва могут получиться только в том случае, если появится выражение $a/0$ или 0^0. Этим доказывается утверждение, сделанное в конце п. 13.

Таким образом, если $f(x)$ — элементарная функция, то $\lim\limits_{x\to x_0} f(x)$ просто равен $f(x_0)$, если в последнем выражении отсутствуют «опасные» выражения вида $a/0$, $\ln 0$ и 0^0; например,

$$\lim_{x\to 1}\frac{\ln(1+\sin x)}{x^2-2x}=\frac{\ln(1+\sin 1)}{1^2-2\cdot 1}=$$
$$=-\ln(1+\sin 1)=-\ln 1{,}8415=-0{,}6105$$

(значения синуса и логарифма взяты из таблиц). Это правило для нахождения предела остается в силе и при манипуляциях с бесконечностями, если только после подстановки предельного значения

аргумента не получается неопределенностей вида $\frac{0}{0}$, $\infty-\infty$, $\frac{\infty}{\infty}$, $0\cdot\infty$ (о которых говорилось в п. 5), 0^0, 1^∞, ∞^0 (о которых будет сказано в п. 15.1), а также выражений вида $\sin\infty$. Например,

$$\lim_{x\to+0}\left(\frac{\ln x}{x}+\cos x\right)=\frac{\ln(+0)}{+0}+\cos(+0)=\frac{-\infty}{+0}+1=-\infty;$$

$$\lim_{x\to+0}\left(x^{\frac{1}{x}}+x^{-\frac{1}{x}}\right)=(+0)^{+\infty}+(+0)^{-\infty}=0+\infty=\infty.$$

О неопределенностях см. § 3 и § IV.4.

3. Как указывалось в п. I 16, график непрерывной функции $y=f(x)$, заданной на интервале $a\leqslant x\leqslant b$, состоит из одного куска. Рассмотрение этого графика (см., например, рис. 105) показывает,

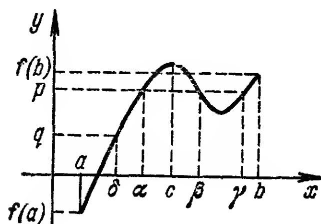


Рис. 105.

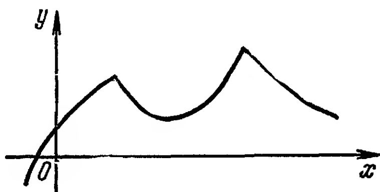


Рис. 106.

что *непрерывная функция, заданная на конечном интервале (включая концы), ограничена и достигает наименьшего в алгебраическом смысле* (на чертеже при $x=a$) *и наибольшего* (при $x=c$) *значений*. Кроме того, *она принимает все промежуточные между $f(a)$ и $f(b)$ значения по крайней мере по одному разу*; так, на чертеже значение $y=q$ принимается один раз — при $x=\delta$, а значение $y=p$ принимается три раза — при $x=\alpha$, β и γ. *Если она положительна при некотором значении $x=x_0$, то она положительна и при всех x, достаточно близких к x_0*. Наконец, из рис. 25 и 26 видно, что если *непрерывная функция монотонна, то обратная к ней функция также непрерывна* (и монотонна).

Мы ограничимся высказанными наглядными соображениями по поводу этих свойств. Строгий их вывод в общем случае оказывается не простым; его можно найти, например, в [48].

Отметим, что непрерывная функция не обязана быть *гладкой*, т. е. иметь график со всюду существующей касательной, как на рис. 105. Она может быть *кусочно-гладкой*, как на рис. 106, так что график ее будет состоять из нескольких гладких дуг и иметь

изломы. Возможны и еще более сложные случаи, которые в нашем курсе не встретятся. Подробней об этом будет сказано в п. IV.3.

15. Некоторые приложения. 1. *Переход к пределу в степени.* Пусть $x \to a$, $y \to b$, причем $x > 0$. Представим x^y в виде $x^y = (e^{\ln x})^y = e^{y \ln x}$. Но $\ln x \to \ln a$ в силу непрерывности логарифмической функции, откуда $y \ln x \to b \ln a$ (предел произведения) и $\exp(y \ln x) \to \exp(b \ln a)$ (непрерывность показательной функции). Итак, $x^y \to e^{b \ln a} = (e^{\ln a})^b = a^b$; другими словами,

$$\lim(x^y) = (\lim x)^{\lim y},$$

т. е. *в степени можно переходить к пределу.* Исключение составляет случай, когда произведение $b \ln a$ будет неопределенностью, т. е. будет иметь вид $0 \cdot \infty$. Это может получиться, если

$$\begin{array}{llll} \ln a = 0; & b = \infty, & \text{т. е.} \quad a = 1 & \text{и} \quad a^b = 1^\infty; \\ \ln a = \infty; & b = 0, & \text{т. е.} \quad a = \infty & \text{и} \quad a^b = \infty^0; \\ \ln a = -\infty; & b = 0, & \text{т. е.} \quad a = 0 & \text{и} \quad a^b = 0^0. \end{array}$$

Получаются как раз три типа неопределенностей, о которых говорилось в п. 14.

Иногда кажется, что должно быть $1^\infty = 1$, так как «единица в любой степени равна единице». Однако 1^∞ — это не единица в некоторой определенной конечной степени, а сокращенный знак для предела выражения вида x^y, где $x \to 1$, а $y \to \infty$. Допустим, например, что $x \to 1 + 0$, т. е. $x > 1$. Тогда из-за изменения x выражение x^y «хочет» стремиться к 1, а из-за изменения y оно «хочет» стремиться к бесконечности (так как если число, большее единицы, возводить в безгранично возрастающую степень, то получится бесконечно большая величина). Получается, что это выражение «тянут» в разные стороны и поэтому в разных примерах может получиться разный результат в зависимости от того, что перетянет Так, в примере (13), где при непосредственной подстановке $h = 0$ получается как раз 1^∞, предел оказался равным e, т. е. «тянущие силы» оказались в каком-то смысле равными. Аналогичное заключение можно сделать о других неопределенностях.

2. *Решение неравенств.* Пусть функция $f(x)$ рассматривается на некотором интервале (a, b) (в частности, может быть, на всей оси x) и требуется решить неравенство

$$f(x) > 0, \qquad (17)$$

т. е найти все те значения x, для которых оно справедливо. Геометрически это означает, что требуется найти участки оси x, для которых график функции $y = f(x)$ проходит выше оси x (на рис. 107 эти участки заштрихованы); но надо иметь в виду, что если дана только функция $f(x)$, то график еще неизвестен.

Решение этой задачи таково. На интервал (a, b) наносят все точки, в которых функция f обращается в нуль или имеет разрыв (на рис. 107 четыре таких точки — три нуля и одна точка разрыва). Этими точками интервал разбивается на несколько частей (на рис. 107 — пять частей), на каждой из которых функция f непрерывна, так как все точки разрыва нанесены, и не обращается в нуль, так как все нули нанесены. Но тогда на каждой из этих частей функция сохраняет знак (см. свойство 3 п. 14); чтобы узнать, каков этот знак, достаточно найти знак функции в какой угодно точке, взятой в рассматриваемой части интервала. После этого, выбрав все части интервала, в которых функция положительна, получим решение неравенства (17).

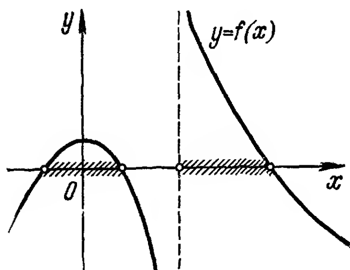


Рис. 107.

Пример. Решим неравенство $\frac{x^3+3x^2-4}{x^2-3} > 0$.

Числитель обращается в нуль при $x=1$ и потому делится на $x-1$. Отсюда

$$y=\frac{x^3+3x^2-4}{x^2-4}=\frac{(x-1)(x^2+4x+4)}{x^2-3}=\frac{(x-1)(x+2)^2}{x^2-3}$$

должно быть > 0. Таким образом, рассматриваемая функция, определенная на всей оси, имеет два нуля ($x=1$ и $x=-2$) и две точки разрыва ($x=\pm\sqrt{3}$). Этими четырьмя точками ось x разбивается на пять частей (рис. 108). Берем

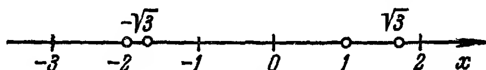


Рис. 108.

в каждой из них по точке и, подставляя в последнюю дробь, находим ее знак (численное значение несущественно); получим

x	-3	$-1,9$	0	$1,1$	2
y	$-$	$-$	$+$	$-$	$+$

Итак, решение рассматриваемого неравенства — это два интервала

$$-\sqrt{3} < x < 1 \quad \text{и} \quad \sqrt{3} < x < \infty.$$

ПРОИЗВОДНЫЕ, ДИФФЕРЕНЦИАЛЫ, ИССЛЕДОВАНИЕ ИЗМЕНЕНИЯ ФУНКЦИИ

§ 1. Производная

1. Примеры, приводящие к понятию производной. К одному из важнейших в математике понятий, понятию производной, мы приходим при изучении скорости изменения функции.

Рассмотрим, например, понятие мгновенной скорости прямолинейного движения точки (*в физическом рассмотрении точкой называется тело, размерами которого мы в данном рассмотрении пренебрегаем*; в разных рассмотрениях точкой может считаться частица вещества, или самолет, или даже небесное тело). Пусть точка движется по оси s слева направо, причем неравномерно, с переменной скоростью. Тогда закон движения математически выражается зависимостью координаты s от времени: $s=f(t)$. Так как скорость переменная, то отношение пройденного пути к истекшему времени дает только *среднюю скорость*. Что же касается мгновенной скорости (скорости в данный момент времени), то она получается следующим образом. Пусть в некоторый момент t движущаяся точка занимает положение A (рис. 109), а через время Δt (по поводу этого обозначения см. п. 1.22) перейдет в положение B, пройдя путь Δs; таким образом,

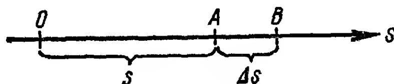


Рис. 109.

$$s=f(t), \quad s+\Delta s=f(t+\Delta t),$$

т. е. $\Delta s=f(t+\Delta t)-f(t)$. Тогда отношение $v_{\text{ср}}=\dfrac{\Delta s}{\Delta t}$ (т. е. пройденный путь, отнесенный к единице истекшего времени) дает среднюю скорость движения за промежуток времени от t до $t+\Delta t$. *Мгновенная* же *скорость* движения в момент t *получается как предел средней скорости в процессе безграничного уменьшения промежутка времени* Δt, т. е.

$$v_{\text{мгн}}=\lim_{\Delta t\to 0} v_{\text{ср}}=\lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t}=\lim_{\Delta t\to 0}\frac{f(t+\Delta t)-f(t)}{\Delta t}. \qquad (1)$$

Говорят также, что мгновенная скорость — это средняя скорость за бесконечно малый промежуток времени (за «*элемент*» времени) или что мгновенная скорость — это отношение бесконечно малого

пути к соответствующему бесконечно малому промежутку времени (т. е. бесконечно малый путь, отнесенный к единице истекшего времени). Оба эти определения являются сокращенными формулировками развернутого определения (1).

Скорость процесса не всегда измеряется пройденным путем, отнесенным к единице истекшего времени. Рассмотрим, например, процесс наполнения сосуда; в данном случае законом наполнения служит зависимость $V=f(t)$ уже наполненного объема от времени. Средней скоростью наполнения за промежуток времени от t до $t+\Delta t$ служит отношение

$$w_{\text{ср}}=\frac{\Delta V}{\Delta t}=\frac{f(t+\Delta t)-f(t)}{\Delta t},$$

тогда как мгновенной скоростью наполнения в момент t — предел

$$w_{\text{мгн}}=\lim_{\Delta t\to 0} w_{\text{ср}}=\lim_{\Delta t\to 0}\frac{\Delta V}{\Delta t}=\lim_{\Delta t\to 0}\frac{f(t+\Delta t)-f(t)}{\Delta t}. \qquad (2)$$

Как видим, с формальной точки зрения получается такое же выражение, как в (1).

Но скорость можно понимать в еще более широком плане, относя изменение некоторой величины не к единице времени, а к единице какой-либо другой величины. Рассмотрим, например, понятие линейной плотности *материальной линии*, т. е. тела, размеры которого учитываются лишь в одном протяжении; поперечным сечением тела мы пренебрегаем, но массой не пренебрегаем. Если эта линия (нить) однородная, то линейная плотность измеряется отношением массы нити к ее длине. Если же нить неоднородная, то ее линейная плотность в разных точках различная. Будем отсчитывать расстояние от одного из концов нити (рис. 110), и пусть масса нити, отвечающая пройденному пути s, равна $M=f(s)$. Если теперь пройден дополнительный путь Δs, то отношение

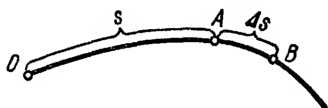


Рис. 110.

$$\rho_{\text{ср}}=\frac{\Delta M}{\Delta s}=\frac{f(s+\Delta s)-f(s)}{\Delta s}$$

представляет собой среднюю линейную плотность нити на участке AB. Предел же

$$\rho=\lim_{\Delta s\to 0}\rho_{\text{ср}}=\lim_{\Delta s\to 0}\frac{\Delta M}{\Delta s}=\lim_{\Delta s\to 0}\frac{f(s+\Delta s)-f(s)}{\Delta s} \qquad (3)$$

представляет собой *линейную плотность нити в точке* (именно, в точке A). Можно сказать, что это — скорость изменения массы нити, отнесенная к единице пройденного пути.

2. Определение производной. Все выражения (1), (2), (3) с математической точки зрения имеют одинаковую структуру и дают

основание для следующего определения. Пусть дана функция $y=f(x)$. Тогда скорость ее изменения, отнесенная к единице изменения аргумента, при значении аргумента x равна

$$y'=\lim_{\Delta s\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}.$$

Эта скорость называется *производной, взятой от переменной y по переменной x*; другими словами, *производная — это предел отношения приращения функции к приращению аргумента, вычисленный в процессе, когда приращение аргумента стремится к нулю*. Так как при различных значениях x указанная скорость, вообще говоря, различна, то производная представляет собой новую функцию x, которую принято обозначать штрихом: $y'=f'(x)$.

Таким образом, в примерах п. 1 скорость движения — это производная от пройденного пути по времени, $v=s'_t$ (нижний индекс указывает на переменную, по которой берется производная) и т. д.

Подсчитаем, например, производную от функции $y=ax^2$. Придавая аргументу приращение Δx, получим новое значение аргумента $x+\Delta x$ и новое значение функции $y+\Delta y=a(x+\Delta x)^2$, так как в выражение функции надо вместо x подставить $x+\Delta x$. Таким образом,

$$\Delta y=a(x+\Delta x)^2-ax^2=2ax\,\Delta x+a(\Delta x)^2.$$

Отсюда

$$y'=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{2ax\,\Delta x+a(\Delta x)^2}{\Delta x}=\lim_{\Delta x\to 0}(2ax+a\,\Delta x)=2ax;$$

отметим, что в последнем предельном переходе изменялось $\Delta x\to 0$, а x считался зафиксированным. Полученный результат можно записать так: $(ax^2)'=2ax$.

Мы предоставляем читателю убедиться аналогичным образом в том, что $(ax^3)'=3ax^2$, $(ax)'=a$ и т. п. Отметим особо, что $x'=1$.

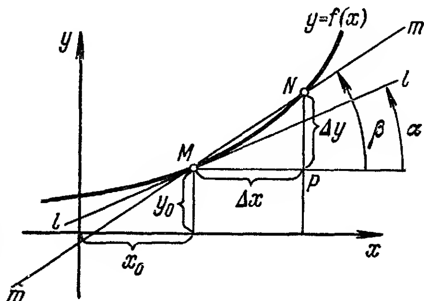


Рис. 111.

3. Геометрический смысл производной. Рассмотрим график функции $f(x)$ (рис. 111). Видно, что $\Delta y/\Delta x=PN/MP=\operatorname{tg}\beta$, т. е. это отношение равно угловому коэффициенту секущей mm. Если $\Delta x\to 0$, то секущая, поворачиваясь вокруг точки M, в пределе переходит в касательную ll, так как *касательная является предельным положением секущей, когда точки пересечения сливаются.* (Это наглядное свойство, которым мы уже пользовались, на самом

деле является просто определением касательной.) Таким образом,

$$y_0' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim \operatorname{tg} \beta = \operatorname{tg} \alpha \qquad (y_0' = f'(x_0)), \tag{4}$$

т. е. *геометрический смысл производной состоит в том, что она равна угловому коэффициенту касательной.* На основании формулы (II.21) легко написать и уравнение касательной ll:

$$y - y_0 = y_0'(x - x_0), \tag{5}$$

где x_0, y_0 — координаты точки касания, а x, y — текущие координаты точки касательной прямой.

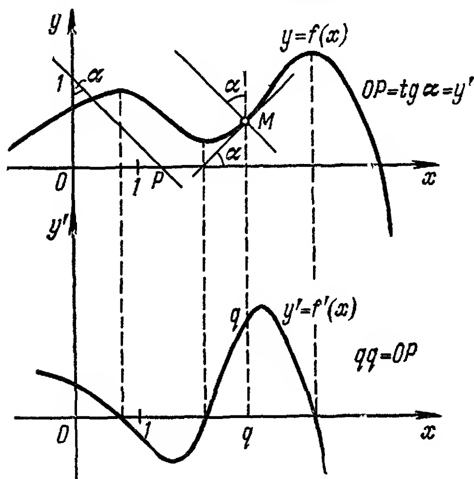


Рис. 112.

Подобным образом уравнение *нормали* к кривой, т. е. перпендикуляра к касательной в точке касания, имеет вид $y - y_0 = -\frac{1}{y_0'}(x - x_0)$ (см. задачу 5 п. II.9).

Умение находить касательную дает возможность с помощью формулы (II.23) находить *угол между линиями в точке их взаимного пересечения*, так как в п. I.26 уже было сказано, что этот угол измеряется углом между касательными к данным линиям в этой точке. Отметим, что этот угол может получиться нулевым, *если линии касаются друг друга*, т. е. если указанные касательные совпадают.

Геометрический смысл производной дает возможность, если дан график функции $y = f(x)$, проследить за наклоном касательной к нему и сразу построить ориентировочный график производной (рис. 112). Для более точного «*графического вычисления производной*» надо проводить касательные к заданному графику и измерять их угловые коэффициенты. На практике оказывается более точным с помощью блестящей (металлической) линейки проводить нормали к графику и измерять их угловые коэффициенты по отношению к оси y. что равносильно. На рис. 112 показано одно из таких построений. Линейка прикладывается перпендикулярно к плоскости графика к некоторой его точке M и поворачивается так, чтобы отражение графика в линейке продолжало сам график без излома в M; в таком положении линейка пойдет как раз по нормали к графику в M. Затем через точку (0, 1) проводится прямая, параллельная построенной нормали, и получающийся отрезок OP переносится в положение qq. Проведя несколько таких построений, можно довольно точно изобразить график производной.

При выводе геометрического смысла (4) производной мы пользовались тем, что переменные x и y — безразмерные, а единицы масштаба по обеим осям — одинаковые. На практике это не всегда так; в общем случае из фор-

мулы (1.10) следует, что $y'_0 = \frac{l_x}{l_y} \operatorname{tg} \alpha$, однако и тогда производная равна угловому коэффициенту касательной.

Отметим, что если производная y' при некотором значении x обращается в бесконечность (при $x = x_1$ на рис. 113), то в соответствующей точке графика касательная имеет угловой коэффициент, равный бесконечности, т. е. параллельна оси y; если производная претерпевает скачок, то и касательная поворачивается скачком, т. е. график имеет излом (при $x = x_2$); если же функция уходит в бесконечность, то и производная уходит в бесконечность (при $x = x_3$).

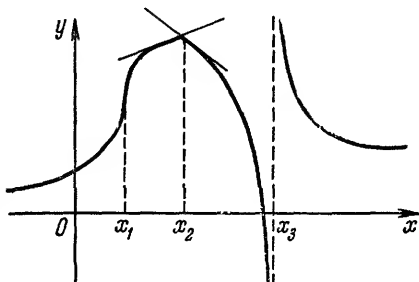


Рис. 113.

4. Основные свойства производной. 1. *Производная постоянной равна нулю*, так как скорость в положении покоя равна нулю. Формальное доказательство: если $y = C = \text{const}$, то

$$\Delta y = C - C = 0, \quad \frac{\Delta y}{\Delta x} = 0,$$

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim 0 = 0.$$

2. *Производная суммы равна сумме производных*. Действительно, если $y(x) = u(x) + v(x)$, то $y(x + \Delta x) = u(x + \Delta x) + v(x + \Delta x)$ и

$$\Delta y = y(x + \Delta x) - y(x) = [u(x + \Delta x) + v(x + \Delta x)] - [u(x) + v(x)] = [u(x) + \Delta u + v(x) + \Delta v] - [u(x) + v(x)] = \Delta u + \Delta v,$$

т. е. *приращение суммы равно сумме приращений*: $\Delta(u + v) = \Delta u + \Delta v$. Отсюда

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim \frac{\Delta u + \Delta v}{\Delta x} = \lim \left(\frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta x} \right) = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} + \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = u' + v'$$

(мы воспользовались тем, что предел суммы равен сумме пределов), что и требуется. Иначе это свойство можно записать так:

$$(u + v)' = u' + v'.$$

Мы рассмотрели сумму двух слагаемых. Ясно, что то же будет при любом числе слагаемых. Аналогичным образом *приращение*

разности равно разности приращений, а производная разности равна разности производных.

Пример. $(x^3-3x^2+x+5)'=(x^3)'-(3x^2)'+(x)'+(5)'=$ $=3x^2-6x+1+0=3x^2-6x+1$ (см. конец п. 2).

3. *Постоянный множитель можно вынести за знак производной,* т. е. $(Cu)'=Cu'$ (где $C=\text{const}$). Действительно, если $y=Cu$, то

$$\Delta y=y(x+\Delta x)-y(x)=Cu(x+\Delta x)-Cu(x)=$$
$$=C[u(x+\Delta x)-u(x)]=C\Delta u;$$

другими словами, *если функцию умножить на константу, то приращение умножится на ту же константу*: $\Delta(Cu)=C\,\Delta u$.

Отсюда $y'=\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\frac{C\,\Delta u}{\Delta x}=C\lim\limits_{\Delta x\to 0}\frac{\Delta u}{\Delta x}=Cu'$.

4. *Формула для производной произведения.* Пусть $y=uv$. Тогда $\Delta y=(u+\Delta u)(v+\Delta v)-uv=(\Delta u)v+u\Delta v+\Delta u\,\Delta v$, откуда

$$y'=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim\frac{(\Delta u)v}{\Delta x}+\lim\frac{u\Delta v}{\Delta x}+\lim\frac{\Delta u\;\Delta v}{\Delta x}=$$
$$=\lim\frac{\Delta u}{\Delta x}v+\lim u\frac{\Delta v}{\Delta x}+\lim\frac{\Delta u}{\Delta x}\frac{\Delta v}{\Delta x}\Delta x=u'v+uv'+u'v'\cdot 0.$$

Итак,

$$(uv)'=u'v+uv'. \qquad (6)$$

Например,

$$[(3x^2+5x)(4x^2-6)]'=(3x^2+5x)'(4x^2-6)+(3x^2+5x)(4x^2-6)'=$$
$$=(6x+5)(4x^2-6)+(3x^2+5x)\,8x=48x^3+60x^2-36x-30.$$

Из формулы (6) легко вывести формулу для производной от произведения нескольких функций. Например, $(uvw)'=[(uv)w]'=$ $=(uv)'w+(uv)w'=(u'v+uv')w+uvw'=u'vw+uv'w+uvw'$. Аналогичный вид имеет формула для произведения любого числа множителей. Отметим, что свойство 3 легко вывести из формулы (6), положив $v=C$.

5. *Формула для производной частного.* Пусть $y=\frac{u}{v}$. Тогда

$$\Delta y=\frac{u+\Delta u}{v+\Delta v}-\frac{u}{v}=\frac{\Delta u\,v-u\Delta v}{v(v+\Delta v)}.$$

Отсюда, представляя в знаменателе Δv в виде $\frac{\Delta v}{\Delta x}\Delta x$, получим

$$y'=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{\frac{\Delta u}{\Delta x}v-u\frac{\Delta v}{\Delta x}}{v\left(v+\frac{\Delta v}{\Delta x}\Delta x\right)}=\frac{u'v-uv'}{v(v+v'\cdot 0)}.$$

Итак,

$$\left(\frac{u}{v}\right)'=\frac{u'v-uv'}{v^2}. \qquad (7)$$

Например,

$$\left(\frac{5x^2}{3x^2+4}\right)'=\frac{(5x^2)'(3x^2+4)-(5x^2)(3x^2+4)'}{(3x^2+4)^2}=$$
$$=\frac{10x(3x^2+4)-5x^2\cdot 6x}{(3x^2+4)^2}=\frac{40x}{(3x^2+4)^2}.$$

6. *Производная сложной функции.* Пусть $y=f(u)$, $u=\varphi(x)$, т. е. y является сложной функцией x. Если x получит приращение Δx, то промежуточная переменная u получит приращение Δu, а поэтому y также получит некоторое приращение Δy. При этом

$$\frac{\Delta y}{\Delta x}=\frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x}. \tag{8}$$

Пусть теперь $\Delta x\longrightarrow 0$. Тогда $\frac{\Delta u}{\Delta x}\longrightarrow u'_x$, а потому $\Delta u=\frac{\Delta u}{\Delta x}\Delta x\longrightarrow$ $\longrightarrow u'_x\cdot 0=0$ и, значит, $\frac{\Delta y}{\Delta u}\longrightarrow y'_u$. Переходя в формуле (8) к пределу, получим

$$y'_x=y'_u u'_x. \tag{9}$$

Эту формулу можно записать так:

$$[f(\varphi(x))]'=f'(\varphi(x))\,\varphi'(x). \tag{10}$$

Аналогично вычисляется производная в случае зависимостей с большим числом промежуточных ступеней. Так, если $y=y(u)$, $u=u(v)$, $v=v(x)$, то $y'_x=y'_u u'_v v'_x$.

Пусть, например, $y=(x^2-5x+3)^3$. Тогда можно обозначить $y=u^3$, где $u=x^2-5x+3$, и по формуле (9) имеем

$$y'_x=y'_u u'_x=(u^3)'_u(x^2-5x+3)'_x=3u^2(2x-5)=$$
$$=3(x^2-5x+3)^2(2x-5);$$

это, конечно, проще, чем раскрывать скобки! На практике не пишут так подробно, а оформляют вычисления так: $[(x^2-5x+3)^3]'=$ $=3(x^2-5x+3)^2(x^2-5x+3)'=3(x^2-5x+3)^2(2x-5)$, т. е., по существу, пользуются формулой (10). При некотором навыке ответ можно писать непосредственно, без промежуточных преобразований. Для этого формулы для производных иногда запоминают в виде $(u^2)'=2uu'$, $(u^3)'=3u^2u'$ (производная берется по x) и т. п.

7. *Производная обратной функции.* Пусть равенство $y=y(x)$ определяет обратную зависимость $x=x(y)$ (см. п. 1.21), для которой мы можем найти производную x'_y. Тогда легко найти производную и от исходной функции. Действительно, $\frac{\Delta y}{\Delta x}=\frac{1}{\Delta x/\Delta y}$, откуда при $\Delta x\longrightarrow 0$ и $\Delta y\longrightarrow 0$.

$$y'_x=\frac{1}{x'_y}. \tag{11}$$

Пусть, например, $y=\sqrt[3]{x}$, откуда $x=y^3$. Тогда

$$y'_x=\frac{1}{x'_y}=\frac{1}{(y^3)'_y}=\frac{1}{3y^2}=\frac{1}{3\sqrt[3]{x^2}}.$$

8. *Производная неявной функции.* Если задана неявная функция в форме $F(x, y)=0$ (см. п. I.20), то для вычисления производной y'_x нужно просто приравнять производные от левой и правой частей заданного соотношения, имея при этом в виду, что y есть функция от x, обращающая это соотношение в тождество. Вообще *непосредственно приравнивать производные от обеих частей равенства можно только в том случае, когда это равенство является тождеством* (*а не уравнением*!).

Пусть, например, задано соотношение

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{12}$$

Тогда

$$\left(\frac{x^2}{a^2}\right)'_x+\left(\frac{y^2}{b^2}\right)'_x=(1)'_x, \quad \text{т. е.} \quad \frac{2x}{a^2}+\frac{2yy'}{b^2}=0; \tag{13}$$

при вычислении производной от второго слагаемого пользовались свойством 6: $\left(\frac{y^2}{b^2}\right)'_x=\frac{1}{b^2}(y^2)'_y\cdot y'_x=\frac{1}{b^2}2yy'$. Из (13) получаем

$$y'=-\frac{b^2x}{a^2y}. \tag{14}$$

5. Производные основных элементарных функций. 1. *Производная синуса.* Пусть $y=\sin x$. Если аргумент изменится и станет равным $x+\Delta x$, то функция станет равной $\sin(x+\Delta x)$. Отсюда

$$\Delta y=\sin(x+\Delta x)-\sin x=2\sin\frac{\Delta x}{2}\cos\left(x+\frac{\Delta x}{2}\right);$$

$$y'=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{\left[2\sin\frac{\Delta x}{2}\cos\left(x+\frac{\Delta x}{2}\right)\right]}{\Delta x}=$$

$$=\lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}\lim_{\Delta x\to 0}\cos\left(x+\frac{\Delta x}{2}\right)=1\cdot\cos x$$

(см. формулу (III.11)). Итак,

$$(\sin x)'=\cos x. \tag{15}$$

2. Предоставляем читателю проверить аналогичным образом, что

$$(\cos x)'=-\sin x. \tag{16}$$

3. *Производную тангенса* вычисляем по формуле (7)

$$(\operatorname{tg} x)'=\left(\frac{\sin x}{\cos x}\right)'=\frac{(\sin x)'\cos x-\sin x(\cos x)'}{\cos^2 x}=$$

$$=\frac{\cos x\cos x-\sin x(-\sin x)}{\cos^2 x}=\frac{1}{\cos^2 x}.$$

4. Аналогично проверяется, что $(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}$.

5. *Производная арксинуса.* Пусть $y = \arcsin x$. Тогда $x = \sin y$, и по формуле (11)

$$y'_x = \frac{1}{x'_y} = \frac{1}{(\sin y)'_y} = \frac{1}{\cos y} = \frac{1}{\pm\sqrt{1-\sin^2 y}} = \frac{1}{\sqrt{1-x^2}};$$

перед радикалом взят $+$, так как значение $\arcsin x$, как известно, берется в 1 и -1-й четвертях, где $\cos y > 0$. Итак,

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}.$$

6. Аналогично проверяется, что

$$(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}; \qquad (17)$$

при этом используется то, что $0 \leqslant \arccos x \leqslant \pi$. Сходство двух последних результатов объясняется формулой

$$\arcsin x + \arccos x \equiv \frac{\pi}{2}, \qquad (18)$$

получаемой следующим образом. Обозначим $\sin\alpha = x$, тогда

$$\cos\left(\frac{\pi}{2} - \alpha\right) = x;$$

из этих формул получаем, что $\alpha = \arcsin x$, $\frac{\pi}{2} - \alpha = \arccos x$. Складывая результаты, приходим к (18).

7. *Производная арктангенса.* Пусть $y = \operatorname{arctg} x$. Тогда $x = \operatorname{tg} y$ и, применяя формулы для производной обратной функции и для производной тангенса, получим

$$y'_x = \frac{1}{x'_y} = \frac{1}{\frac{1}{\cos^2 y}} = \cos^2 y = \frac{1}{1+\operatorname{tg}^2 y} = \frac{1}{1+x^2}.$$

Итак,

$$(\operatorname{arctg} x)' = \frac{1}{1+x^2}.$$

8. *Производная логарифмической функции.* Пусть $y = \ln x$. Тогда в силу формулы (III.12), в которой надо положить $h = \frac{\Delta x}{x}$,

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\ln(x+\Delta x) - \ln x}{\Delta x} = \lim_{\Delta x \to 0} \frac{\ln\left(1+\frac{\Delta x}{x}\right)}{\frac{\Delta x}{x}x} = \frac{1}{x}.$$

Итак,

$$(\ln x)' = \frac{1}{x}.$$

Применяя формулу (I.14) и учитывая, что $\ln a = \text{const}$, получим

$$(\log_a x)' = \left(\frac{\ln x}{\ln a}\right)' = \frac{1}{\ln a}(\ln x)' = \frac{1}{x \ln a}.$$

9. *Производная показательной функции.* Если $y = a^x$, то

$$x = \log_a y \quad \text{и} \quad y'_x = \frac{1}{x'_y} = \frac{1}{\frac{1}{y \ln a}} = y \ln a = a^x \ln a.$$

Итак,

$$(a^x)' = a^x \ln a;$$

в частности, $(e^x)' = e^x$.

10. *Производная степенной функции.* Согласно формуле для производной сложной функции

$$(x^n)' = [(e^{\ln x})^n]' = (e^{n \ln x})' = e^{n \ln x} n \frac{1}{x} = x^n n \frac{1}{x} = nx^{n-1}.$$

Итак,

$$(x^n)' = nx^{n-1}.$$

Эта формула справедлива для любого n, целого или нецелого. Например, $(\sqrt{x})' = \left(x^{\frac{1}{2}}\right)' = \frac{1}{2} x^{\frac{1}{2}-1} = \frac{1}{2\sqrt{x}}$; $\left(\frac{1}{x}\right)' = (x^{-1})' = -1x^{-1-1} = \frac{-1}{x^2}$ и т. п.

11. *Производные гиперболических функций* (см. п. I.28). Имеем

$$(\operatorname{sh} x)' = \left(\frac{e^x - e^{-x}}{2}\right)' = \frac{e^x - e^{-x}(-x)'}{2} = \frac{e^x + e^{-x}}{2} = \operatorname{ch} x;$$

аналогично

$$(\operatorname{ch} x)' = \operatorname{sh} x;$$

$$(\operatorname{th} x)' = \left(\frac{\operatorname{sh} x}{\operatorname{ch} x}\right)' = \frac{(\operatorname{sh} x)' \operatorname{ch} x - \operatorname{sh} x (\operatorname{ch} x)'}{\operatorname{ch}^2 x} = \frac{\operatorname{ch}^2 x - \operatorname{sh}^2 x}{\operatorname{ch}^2 x} = \frac{1}{\operatorname{ch}^2 x};$$

$$(\operatorname{arsh} x)' = [\ln(x + \sqrt{x^2+1})]' = \frac{1}{x+\sqrt{x^2+1}}(x + \sqrt{x^2+1})' =$$

$$= \frac{1}{x+\sqrt{x^2+1}}\left(1 + \frac{(x^2+1)'}{2\sqrt{x^2+1}}\right) = \frac{1}{x+\sqrt{x^2+1}}\left(1 + \frac{2x}{2\sqrt{x^2+1}}\right) =$$

$$= \frac{1}{x+\sqrt{x^2+1}} \frac{\sqrt{x^2+1}+x}{\sqrt{x^2+1}} = \frac{1}{\sqrt{x^2+1}}.$$

И в этих формулах обнаруживается значительная аналогия с тригонометрическими функциями.

12. Выведенные формулы (*табличные производные*) нужно знать наизусть, так как они будут применяться систематически. С их помощью, пользуясь правилами п. 4, можно вывести производную любой элементарной функции. Например,

$$\left(\sqrt[3]{x}\, 2^{\operatorname{tg} 5x}\right)' = \left(\sqrt[3]{x}\right)' 2^{\operatorname{tg} 5x} + \sqrt[3]{x}\left(2^{\operatorname{tg} 5x}\right)';$$

но

$$(\sqrt[3]{x})' = \left(x^{\frac{1}{3}}\right)' = \frac{1}{3} x^{\frac{1}{3}-1} = \frac{1}{3\sqrt[3]{x^2}},$$

а по формуле для производной сложной функции

$$(2^{\operatorname{tg} 5x})' = 2^{\operatorname{tg} 5x} \ln 2 \, (\operatorname{tg} 5x)' = 2^{\operatorname{tg} 5x} \ln 2 \frac{1}{\cos^2 5x} (5x)' = 2^{\operatorname{tg} 5x} \ln 2 \frac{1}{\cos^2 5x} \cdot 5.$$

Отсюда после вынесения общего множителя за скобки получим

$$(\sqrt[3]{x}\, 2^{\operatorname{tg} 5x})' = \frac{2^{\operatorname{tg} 5x}}{3\sqrt[3]{x^2}} \left(1 + 15 \ln 2 \frac{x}{\cos^2 5x}\right).$$

После некоторого навыка эти выкладки осуществляются значительно быстрее, без промежуточных преобразований.

13. В некоторых случаях перед вычислением производной полезно *предварительное логарифмирование*. Пусть, например, надо найти производную $(x^{\sin x})'$. Тогда пишем $y = x^{\sin x}$; $\ln y = \sin x \ln x$; $(\ln y)'_x = (\sin x \ln x)'$;

$$\frac{1}{y} y' = \cos x \ln x + \sin x \frac{1}{x};$$

в левой части мы применили формулу для производной сложной функции. Отсюда окончательно

$$y' = (x^{\sin x})' = x^{\sin x} \left(\cos x \ln x + \sin x \frac{1}{x}\right).$$

Этот метод иногда применяется также, если нужно найти производную произведения нескольких множителей: после логарифмирования получается сумма, от которой найти производную бывает легче.

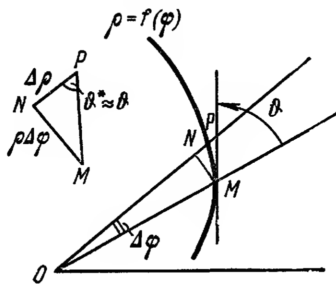


Рис. 114.

6. Касательная в полярных координатах. Задача о нахождении касательной к линии, заданной своим уравнением в декартовых координатах, была решена в п. 3. Пусть теперь линия задана своим уравнением $\rho = f(\varphi)$ в полярных координатах. Конечно, тогда для нахождения касательной можно было бы перейти к декартову уравнению, но проще решать эту задачу прямо в полярных координатах (рис. 114). Придадим φ малое приращение $\Delta\varphi$ и рассмотрим бесконечно малый криволинейный треугольник MNP, образованный двумя координатными линиями и рассматриваемым графиком; для удобства этот треугольник на рис. 114 изображен отдельно. С точностью до малых высшего порядка этот треугольник можно считать «настоящим» (т. е. прямолинейным), причем прямоугольным, так как $\angle N = 90°$ (почему?). Отсюда

$$\operatorname{ctg} \vartheta = \lim \operatorname{ctg} \vartheta^* = \lim \frac{\Delta\rho}{\rho\,\Delta\varphi} = \frac{1}{\rho} \rho'_\varphi. \qquad (19)$$

Пример 1. Для логарифмической спирали (см. п. II.5)

$$\operatorname{ctg}\vartheta=\frac{1}{e^{k\varphi}}(ke^{k\varphi})=k=\text{const},$$

т. е. эта спираль пересекает все координатные лучи под одним и тем же углом. Легко установить связь этого свойства со свойством, указанным в п. II 5: одно свойство выводится из другого.

Пример 2. Рассмотрим полярное уравнение параболы относительно ее фокуса [уравнение (II.29), в котором надо положить $\varepsilon=1$]. По формуле (19)

$$\operatorname{ctg}\vartheta=\frac{1+\cos\varphi}{p}\,\frac{p\sin\varphi}{(1+\cos\varphi)^2}=\frac{\sin\varphi}{1+\cos\varphi}=\frac{2\sin\frac{\varphi}{2}\cos\frac{\varphi}{2}}{2\cos^2\frac{\varphi}{2}}=$$

$$=\operatorname{tg}\frac{\varphi}{2}=\operatorname{ctg}\left(\frac{\pi}{2}-\frac{\varphi}{2}\right);$$

$$\vartheta=\frac{\pi}{2}-\frac{\varphi}{2}.$$

Поэтому если провести через точку M параболы прямую, параллельную оси, то (рис. 115) $\alpha+\vartheta=\pi-\varphi$, т. е.

$$\alpha=\pi-\varphi-\vartheta=\pi-\varphi-\left(\frac{\pi}{2}-\frac{\varphi}{2}\right)=\frac{\pi}{2}-\frac{\varphi}{2}=\vartheta=\beta.$$

Отсюда следует *основное оптическое свойство параболы*: при распространении света в плоскости параболы, если в ее фокусе поместить источник

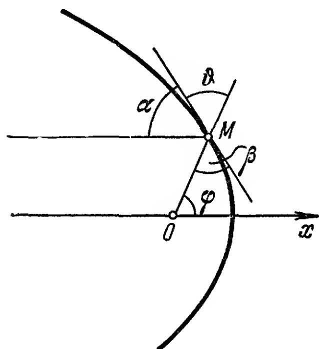


Рис. 115.

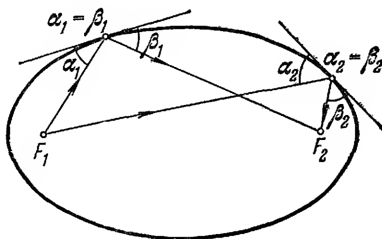


Рис. 116.

света, лучи света после отражения от параболы все пойдут параллельно ее оси. Этим объясняется то, что, например, прожекторам придают форму поверхности, получающейся от вращения параболы вокруг ее оси.

Чуть сложнее выводится аналогичное оптическое свойство эллипса и гиперболы: так, у эллипса лучи света, вышедшие из одного фокуса, после отражения от эллипса все собираются в другом фокусе (рис. 116). Параболу можно рассматривать как эллипс, у которого один из фокусов удален на бесконечность (см. конец п. II.12), поэтому из оптического свойства эллипса в пределе получается оптическое свойство параболы.

§ 2. Дифференциал

7. Физические примеры. Понятие дифференциала, тесно связанное с понятием производной, также является одним из важнейших в математике. Мы его проиллюстрируем на тех же примерах, что были рассмотрены нами в п. 1.

Пусть при прямолинейном движении точки по закону $s=f(t)$ она обладала в некоторый момент t скоростью $v=s'_t=f'(t)$. Если теперь пройдет дополнительное время Δt, то точка пройдет дополнительный путь Δs, который в случае неравномерности движения зависит от Δt сложным образом, так как скорость движения все время меняется. Однако если истекшее время Δt невелико, то и скорость не успеет существенно измениться и потому движение в промежутке времени от t до $t+\Delta t$ является «почти равномерным». В этом случае при подсчете пути не будет большой ошибки, если считать движение равномерным, т. е. происходящим с постоянной скоростью, именно той, которой точка обладала в момент t.

Получающийся при таком подсчете путь равен $v\Delta t=s'_t\Delta t=$ $=f'(t)\Delta t$; он прямо пропорционален истекшему времени Δt, называется *дифференциалом пути* и обозначается ds (этот символ надо понимать как единый, а не как произведение d на s), $ds=s'_t\Delta t$. Конечно, фактический путь Δs отличается от этого «примысленного» пути ds, так как за время Δt, даже малое, скорость все же успевает измениться. Однако в силу сказанного, если этот промежуток времени достаточно мал, можно приближенно считать

$$\Delta s \approx ds, \tag{20}$$

причем с тем бóльшим основанием, чем меньше Δt, так как тем меньше успеет измениться скорость движения. Если же промежуток Δt бесконечно мал, то, как мы увидим в п. 8, Δs и ds отличаются друг от друга на величину высшего порядка малости. Во многих вопросах такими величинами можно пренебрегать; тогда говорят, что *дифференциал пути* — это не что иное, как *бесконечно малый путь*, т. е. путь, пройденный за бесконечно малый промежуток времени. Впрочем, конечно, дифференциал пути может и не быть бесконечно малым, но чем он больше, тем формула (20) менее точна. В то же время вычислять ds как путь при равномерном движении гораздо легче, чем фактический путь Δs; этим объясняется то, что формулой (20) пользуются и при не очень малых Δt.

Таким же образом во втором примере дифференциал объема dV — это тот объем, который наполнился бы, если бы в промежутке времени от t до $t+\Delta t$ скорость наполнения оставалась постоянной, равной скорости в момент t, т. е. $dV=V'_t\Delta t$. В третьем примере дифференциал массы — это масса, которой обладал бы участок AB линии (см. рис. 110), если бы линейная плотность на этом участке

была постоянной, равной плотности в точке A, т. е. $dM = \rho\Delta s = M'_s \Delta s$.

Во всех случаях замена истинного изменения какой-либо величины ее дифференциалом означает переход от неравномерных процессов, неоднородных объектов и т. п. к равномерным, однородным. Эта замена основана на том, что на протяжении малого промежутка времени всякий процесс «почти равномерен», на малой протяженности всякий объект «почти однороден» и т. п.

8. Определение дифференциала и связь его с приращением. Дадим теперь общее определение дифференциала. Пусть $y = f(x)$, и аргумент, первоначально принимавший некоторое значение x, получил приращение Δx. Тогда *дифференциалом функции называется произведение*

$$dy = df(x) = y' \Delta x = f'(x) \Delta x; \tag{21}$$

это — то приращение, которое получила бы функция, если бы она на интервале от x до $x + \Delta x$ изменялась с той же скоростью, что и при значении x аргумента.

Вычисление дифференциала функции называется ее *дифференцированием*; оно очень просто осуществляется по формуле (21). Пусть, например, $y = \sin x$; тогда $dy = (\sin x)' \Delta x = \cos x \Delta x$, т. е. $d \sin x = \cos x \Delta x$.

Аналогично $d \operatorname{tg} x = \frac{1}{\cos^2 x} \Delta x$; $d(x^3) = 3x^2 \Delta x$ и т. п. Таким образом, при дифференцировании функции надо вычислить ее производную, а результат умножить на Δx; поэтому процесс вычисления производной тоже часто называют дифференцированием. Однако *ни в коем случае нельзя путать производную и дифференциал друг с другом*. Производная функции $y = f(x)$ зависит только от x, тогда как дифференциал зависит также от Δx; в приложениях дифференциал обычно считается величиной бесконечно малой, тогда как производная — величиной конечной; если величины x и y размерные, то

$$[dy] = [\Delta y] = [y]; \qquad [y'_x] = \frac{[\Delta y]}{[\Delta x]} = \frac{[y]}{[x]}.$$

Отметим, в частности, что

$$dx = x'_x \Delta x = 1 \, \Delta x = \Delta x,$$

т. е. *дифференциал независимого переменного равен его приращению*. Это дает возможность представить формулу (21) в виде

$$dy = f'(x) \, dx = y' \, dx \tag{22}$$

и, с другой стороны, записать производную в виде отношения дифференциалов:

$$y'_x = \frac{dy}{dx} \quad \text{или, что то же,} \quad f'(x) = \frac{df(x)}{dx}.$$

Геометрический смысл дифференциала функции показан на рис. 117: *он равен приращению ординаты касательной*. Таким образом, замена приращения функции на ее дифференциал геометрически означает, что график функции заменяется отрезком касательной к нему в точке A. Ясно, что для такой замены имеются основания, если Δx достаточно малó.

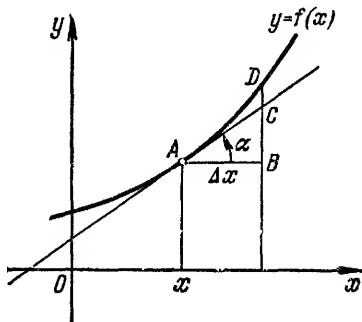


Рис. 117.
$BD=\Delta y$; $BC=AB\operatorname{tg}\alpha=\Delta x\, y'=dy$

Для выяснения связи дифференциала с приращением заметим, что $\frac{\Delta y}{\Delta x}\underset{\Delta x\to 0}{\longrightarrow} y'$, т. е. $\frac{\Delta y}{\Delta x}=y'+\alpha$, где $\alpha\to 0$ при $\Delta x\to 0$. Отсюда

$$\Delta y=y'\,\Delta x+\alpha\,\Delta x=dy+\beta, \quad (23)$$

где $\beta=\alpha\,\Delta x$ — величина высшего порядка малости по сравнению с Δx (на рис. 117 она изображается отрезком CD). Это равенство выражают словами «*дифференциал есть главная линейная часть приращения функции*»: главная — так как он отличается от приращения на величину β высшего порядка, а линейная — потому что он прямо пропорционален Δx (ср. п. I 22). Если $y'\neq 0$, то dy и $dx=\Delta x$ имеют одинаковый порядок малости, а потому β в формуле (23) имеет высший порядок малости, чем dy, т. е. dy и Δy — бесконечно малые эквивалентные (см. п. III.8).

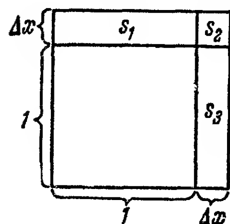


Рис. 118.
$\Delta y=s_1+s_2+s_3$;
$dy=s_1+s_3$; $(\Delta x)^2=s_2$.

Покажем на примере, какая ошибка получается при замене приращения функции ее дифференциалом. Пусть $y=x^2$ и аргумент, принимая первоначальное значение $x=1$, получит приращение Δx. Тогда

$$\Delta y=(1+\Delta x)^2-1^2=2\Delta x+\Delta x^2;$$
$$dy=y'\,\Delta x=2\cdot 1\Delta x=2\Delta x.$$

Таким образом, Δy и dy отличаются друг от друга на величину $(\Delta x)^2$ второго порядка малости (рис. 118). В частности,

при $\Delta x=0{,}1$ $\quad\Delta y=0{,}21$; $\quad dy=0{,}2$; ошибка 5%;
при $\Delta x=0{,}01$ $\quad\Delta y=0{,}0201$; $\quad dy=0{,}02$; ошибка 0,5%;
при $\Delta x=-0{,}001$ $\quad\Delta y=-0{,}001999$; $\quad dy=-0{,}002$; ошибка 0,05% и т. д.

Хорошо видно, что при замене Δy на dy относительная погрешность при уменьшении $|\Delta x|$ быстро уменьшается.

Функция, обладающая дифференциалом, называется *дифференцируемой*. Другими словами, дифференцируемая функция — это функция, малое приращение которой имеет главную линейную часть, т. е. функция, которую на любом малом участке изменения аргумента можно приближенно заменить на линейную (такая замена называется *линеаризацией*). У дифференцируемой функции производная должна

быть конечной, а сама функция должна быть непрерывной при рассматриваемых значениях аргумента, так как из (23) видно, что при бесконечно малом Δx и Δy будет бесконечно малым. В то же время непрерывная функция может быть не всюду дифференцируемой: например, функция, изображенная на рис. 113, перестает быть дифференцируемой не только в точке разрыва $x=x_3$, но и в точках непрерывности $x=x_1$ и $x=x_2$. Выдающийся чешский математик Б. Больца́но (1781—1848) и независимо от него выдающийся немецкий математик К. Ве́йерштрасс (1815—1897) открыли (первый в 1830 г., а второй в 1860 г., так как работа Больцано не была опубликована) существование непрерывных функций, не дифференцируемых *ни при каком* значении аргумента. Долгое время эти функции считались математическим курьезом, но позже обнаружилось, что при описании процессов типа броуновского движения такие функции чрезвычайно важны. В нашем вводном курсе мы не будем даже принимать во внимание возможность подобных монстров.

9. Свойства дифференциала. Так как дифференциал функции получается в результате простого умножения ее производной на дифференциал независимого переменного, то из каждого свойства производной (см. п. 4) легко вывести соответствующее свойство дифференциала. Например, умножая обе части равенства $(u+v)'=$ $=u'+v'$ на dx, получим $(u+v)'\,dx=u'dx+v'dx$ или, что то же самое,

$$d\,(u+v)=du+dv$$

(*дифференциал суммы равен сумме дифференциалов*). Аналогично получаем формулу

$$d\,(uv)=(du)\,v+u\,dv \tag{24}$$

и т. п. В п. IX.12 мы увидим, что эти формулы справедливы и для случая любого числа независимых переменных.

Особенно важный вывод вытекает из формулы для производной сложной функции. Пусть $y=f(x)$ и x сначала является независимой переменной. Тогда для вычисления dy можно пользоваться любой из формул (21) или (22), так как в этом случае $\Delta x=dx$. Пусть теперь x зависит от некоторой третьей величины, например $x=x(t)$. Тогда уже $\Delta x\neq dx$, но оказывается, что *формула* (22) *все равно остается справедливой* (а формула (21), вообще говоря, нарушается). Действительно,

$$dy=y'_t\,dt=y'_x x'_t\,dt=y'_x\,dx,$$

что и требуется. Поэтому при вычислении дифференциалов обычно пользуются формулой (22) (а не (21)), так как она остается справедливой, инвариантной (неизменной) во всех случаях.

Применим это свойство инвариантности для вычисления *производной от функции, заданной параметрически* (см. п. II.6). Пусть

$x=x(t)$; $y=y(t)$ (t — параметр). Тогда $dx=\dot{x}dt$; $dy=\dot{y}dt$ (точкой сверху обычно обозначается производная по параметру), откуда

$$y'=\frac{dy}{dx}=\frac{\dot{y}}{\dot{x}}. \qquad (25)$$

Все эти свойства дифференциалов применяются, в частности, при линеаризации зависимостей между величинами, т. е. при переходе от общей, нелинейной зависимости к линейной зависимости между приращениями этих величин. Такая линеаризация возможна при малых изменениях рассматриваемых величин и основана на отбрасывании величин высшего порядка малости.

Так, например, уравнение (II.30) определяет нелинейную связь между координатами точки $M(x;\ y)$ линии второго порядка. Однако пусть точка M меняется вблизи некоторой фиксированной точки $M_0(x_0;\ y_0)$, т. е. приращения $x-x_0=\xi$, $y-y_0=\eta$ малы. Дифференцируя уравнение (II.30), а затем заменяя дифференциалы приращениями, приходим к линеаризованному уравнению

$$2Ax_0\xi+2By_0\xi+2Bx_0\eta+2Cy_0\eta+D\xi+E\eta=0, \qquad (26)$$

которое определяет линейную зависимость между ξ и η. Так как при выводе уравнения (26) мы заменяли дифференциалы dx, dy на приращения ξ, η, то этому уравнению точка M линии удовлетворяет лишь с точностью до величин высшего порядка малости. Точно уравнению (26) удовлетворяют точки касательной к линии (II.30), проведенной в точке M_0.

Линеаризация широко применяется в физике, в частности, при составлении дифференциальных уравнений (п. XIV.6).

10. Применение дифференциала в приближенных вычислениях. Дифференциал широко применяется в приближенных вычислениях. Прежде всего *довольно часто приращение функции заменяют ее дифференциалом, который обычно вычислять проще.*

Допустим, что дана функция $y=f(x)$, для которой известно некоторое значение $f(a)$; пусть после этого аргумент получил малое приращение $\Delta x=h$. Тогда можно положить

$$f(a+h)-f(a)=\Delta y\approx dy=f'(a)h,$$

т. е.

$$f(a+h)\approx f(a)+f'(a)h. \qquad (27)$$

Выбирая в качестве $f(x)$ конкретные функции $\sqrt[n]{x}$, $\sin x$, $\ln x$ и т. д., получим приближенные формулы

$$\left.\begin{aligned}&\sqrt[n]{a+h}\approx\sqrt[n]{a}+\frac{h}{n\sqrt[n]{a^{n-1}}};\\&\sin(a+h)\approx\sin a+h\cos a;\\&\ln(a+h)\approx\ln a+\frac{h}{a}\end{aligned}\right\} \qquad (28)$$

и т. д., пригодные для достаточно малых $|h|$. О том, как уточнить эти формулы и оценить их погрешность, будет сказано в пп. 15—16.

Пусть, например, мы знаем, что $\ln 2=0{,}693$. Тогда, вычисляя до 0,001,

$$\ln 2{,}1=\ln(2+0{,}1)\approx\ln 2+\frac{0{,}1}{2}=0{,}693+0{,}050=0{,}743;$$

по таблице же оказывается $\ln 2{,}1=0{,}742$, т. е. ошибка меньше 0,2%.

Иногда перед применением формул (28) требуется предварительное преобразование величины, которую нужно вычислить. Например, вычисление $\sqrt[3]{2}$ нехорошо проводить так:

$$\sqrt[3]{2}=\sqrt[3]{1+1}\approx\sqrt[3]{1}+\frac{1}{3\sqrt[3]{1^2}}=1+\frac{1}{3}=1{,}333,$$

так как значение $h=1$ в данном случае вряд ли можно считать малым по сравнению с $a=1$. Здесь удобно положить $\sqrt[3]{2}=\sqrt[3]{2m^3}/m$ и подобрать целое m так, чтобы $2m^3$ оказалось по возможности ближе к какому-либо полному кубу. Можно взять $m=4$, так как $2\cdot 4^3=128$ близко к $125=5^3$; тогда получим, вычисляя до 0,0001,

$$\sqrt[3]{2}=\frac{1}{4}\sqrt[3]{2\cdot 4^3}=\frac{1}{4}\sqrt[3]{128}=\frac{1}{4}\sqrt[3]{125+3}\approx\frac{1}{4}\left(\sqrt[3]{125}+\frac{3}{3\sqrt[3]{125^2}}\right)=$$
$$=\frac{1}{4}\left(5+\frac{1}{25}\right)=\frac{1}{4}\cdot 5{,}0400=1{,}2600.$$

По таблицам же $\sqrt[3]{2}=1{,}2599$, т. е. ошибка меньше 0,01%.

Дифференциалы применяются также при оценке погрешности. Допустим, что величины x и y связаны функциональной зависимостью $y=f(x)$ и известно приближенное значение $\bar{x}$ величины x с предельной абсолютной погрешностью α_x (см. п. 1.7). Тогда в качестве приближенного значения y надо взять, конечно, $\bar{y}=f(\bar{x})$. Для подсчета предельной абсолютной погрешности α_y заметим, что на самом деле $x=\bar{x}+h$, где $|h|<\alpha_x$, откуда, если α_x, а следовательно и h, мало́, то

$$y=\bar{y}+\Delta y\approx\bar{y}+dy=\bar{y}+f'(\bar{x})\,h,$$

т. е.

$$|y-\bar{y}|\approx|f'(\bar{x})|\,|h|<|f'(\bar{x})|\,\alpha_x.$$

Итак, можно положить

$$\alpha_y=|f'(\bar{x})|\,\alpha_x. \tag{29}$$

Пусть, например, $y=x^n$. Тогда

$$\alpha_y=|n|\,\bar{x}^{n-1}\alpha_x,$$

а соответствующие предельные относительные погрешности связаны простой формулой:

$$\delta_y=\frac{\alpha_y}{\bar{y}}=\frac{|n|\,\bar{x}^{n-1}\alpha_x}{\bar{x}^n}=\frac{|n|\,\alpha_x}{\bar{x}}=|n|\,\delta_x.$$

В качестве другого примера рассмотрим величину $\ln 10{,}7$, где значение 10,7 — приближенное, известное с точностью до 0,1. По таблице $\ln 10{,}7=2{,}3702$, однако ясно, что в таком ответе слишком много знаков. Чтобы разобраться, какова точность ответа, заметим, что в данном случае $\alpha_x=0{,}1$, откуда

$$\alpha_y=\frac{1}{10{,}7}\,0{,}1\approx 0{,}01,$$

т. е. ответ надо писать в виде $\ln 10{,}7=2{,}37$.

§ 3. Производные и дифференциалы высших порядков

11. Производные высших порядков. Пусть $y=f(x)$. Тогда производная $y'=f'(x)$, которую мы изучали в § 1, называется *производной первого порядка*. Она в свою очередь является функцией x и потому от нее можно взять производную, которая называется *производной второго порядка* от исходной функции:

$$y''=(y')'=f''(x).$$

Аналогично определяется *производная третьего порядка*

$$y'''=(y'')'=f'''(x).$$

Дальнейшие производные обозначаются $y^{(4)}=y^{\mathrm{IV}}$, $y^{(5)}=y^{\mathrm{V}}$ и т. д. Например, $(x^3)'=3x^2$, $(x^3)''=(3x^2)'=6x$; $(\sin x)'=\cos x$, $(\sin x)''=$ $=(\cos x)'=-\sin x$ и т. п. Производная второго порядка часто имеет непосредственный физический смысл: так, в первом примере п. 1 производная второго порядка от пути по времени — это скорость изменения мгновенной скорости, т. е. мгновенное ускорение. О применении производных высших порядков будет сказано в п. 15 и далее.

Формула для производной суммы очень проста. Если $y=u+v$, то $y'=u'+v'$, $y''=(u'+v')'=u''+v''$ и т. д. Вообще

$$(u+v)^{(n)}=u^{(n)}+v^{(n)}.$$

Что касается формулы для производной произведения, то

$$\begin{aligned}
(uv)' &= u'v+uv';\\
(uv)'' &= (u'v+uv')'=u''v+u'v'+u'v'+uv''=u''v+2u'v'+uv'';\\
(uv)''' &= (u''v+2u'v'+uv'')'=u'''v+2u''v'+u'v''+u''v'+2u'v''+uv'''=\\
&= u'''v+3u''v'+3u'v''+uv''' \text{ и т д.};
\end{aligned} \tag{30}$$

при этом, вычисляя очередную производную, мы сначала во всех членах дифференцируем первый множитель, а затем во всех членах — второй. Эти вычисления идут по той же схеме, как при последовательном разложении выражений $(a+b)^2$, $(a+b)^3$ и т. д.:

$$\begin{aligned}
a+b)^2 &= (a+b)(a+b)=a^2+ab+ab+b^2=a^2+2ab+b^2;\\
(a+b)^3 &= (a^2+2ab+b^2)(a+b)=a^3+2a^2b+ab^2+a^2b+2ab^2+b^3=\\
&= a^3+3a^2b+3ab^2+b^3 \text{ и т. д.}
\end{aligned} \tag{31}$$

Поэтому в формулах (30) получаются такие же коэффициенты, как и в формулах (31). В общем случае можно написать (*формула Лейбница*)

$$(uv)^{(n)}=u^{(n)}v+\binom{n}{1}u^{(n-1)}v'+\binom{n}{2}u^{(n-2)}v''+\ldots+uv^{(n)}, \tag{32}$$

где $\binom{n}{1}$, $\binom{n}{2}$, ... — это так называемые *биномиальные коэффициенты*, т. е. коэффициенты, получающиеся при разложении степени $(a+b)^n$.

Вычисление производной неявной функции мы покажем на примере (12). При дальнейшем дифференцировании равенства (13) получим

$$\frac{2}{a^2}+\frac{2}{b^2}(y'y'+yy'')=0, \text{ т. е. } y''=-\frac{b^2}{y}\left(\frac{1}{a^2}+\frac{y'^2}{b^2}\right)=$$

$$=-\frac{b^2}{y}\left(\frac{1}{a^2}+\frac{b^2x^2}{a^4y^2}\right)=-\frac{b^4}{a^2y^3}\left(\frac{y^2}{b^2}+\frac{x^2}{a^2}\right)=-\frac{b^4}{a^2y^3} \qquad (33)$$

(при этом мы воспользовались выражением (14) для y'). Аналогично вычисляются дальнейшие производные. Формулу (33) можно получить также, дифференцируя равенство (14).

Остановимся еще на дифференцировании функции, заданной параметрически. Дифференцируя формулу (25), получим

$$y''=\frac{d\,(y')}{dx}=\frac{d\left(\frac{\dot y}{\dot x}\right)}{dx}=\frac{\frac{\ddot y\dot x-\dot y\ddot x}{\dot x^2}\,dt}{\dot x dt}=\frac{\ddot y\dot x-\dot y\ddot x}{\dot x^3}$$

(двумя точками обозначается производная второго порядка по параметру). При этом мы, как и при выводе формулы (25), пользуемся тем, что при вычислении дифференциала первого порядка можно не обращать внимания на то, какая переменная является независимой. Подобным образом вычисляются дальнейшие производные.

12. Дифференциалы высших порядков. Пусть $y=f(x)$. Тогда

$$dy=f'(x)\,dx. \qquad (34)$$

Это — *дифференциал первого порядка. Дифференциал второго порядка* d^2y — это дифференциал от дифференциала первого порядка; при этом, если x — независимая переменная, то при таком вторичном дифференцировании dx считается независимым от x и выносится как постоянная величина:

$$d^2y=d\,(dy)=d\,(f'(x)\,dx)=d\,(f'(x))\,dx=$$

$$=(f'(x))'\,dx\,dx=f''(x)\,dx^2, \qquad (35)$$

где принято обозначение $dx^2=(dx)^2$. Аналогично получим

$$d^3y=d\,(d^2y)=f'''(x)\,dx^3 \qquad (36)$$

и т. д. Это дает возможность записать производные высших порядков в виде отношения дифференциалов:

$$y''=\frac{d^2y}{dx^2}, \quad y'''=\frac{d^3y}{dx^3} \quad \text{и т. д.} \qquad (37)$$

Кроме того, видно, что если dy имел первый порядок малости в сравнении с dx, то d^2y имеет второй порядок малости, d^3y — третий и т. д. Далее отметим, что

$$d^2x=d\,(dx)=d\,(1dx)=dxd\,(1)=0;$$

второй дифференциал независимой переменной равен нулю; конечно, нулю равны и дальнейшие дифференциалы независимой переменной.

Если x не является независимой переменной (или нам неизвестно, является или нет), то, как мы видели в п. 8, формула (34) все равно справедлива. Однако при ее дальнейшем дифференцировании dx уже нельзя считать постоянным, а надо пользоваться правилом дифференцирования произведения [формула (24)]:

$$d^2y = d(f'(x)\,dx) = d(f'(x))\,dx + f'(x)\,d(dx) = f''(x)\,dx^2 + f'(x)\,d^2x; \tag{38}$$

аналогично находятся

$$d^3y = f'''(x)\,dx^3 + 3f''(x)\,dx\,d^2x + f'(x)\,d^3x \tag{39}$$

(проверьте!) и дальнейшие дифференциалы. Если теперь окажется, что x — независимая переменная, то $d^2x = d^3x = 0$ и формула (38) переходит в формулу (35), а (39) — в (36). Итак, *формулами* (35)—(37) *можно пользоваться, только если x — независимая переменная.*

Понятия производной и дифференциала и основные правила действий с ними, т. е. правила *дифференциального исчисления*, разработали Ньютон (1666) и Лейбниц (1684), хотя в частных задачах эти понятия применялись и ранее.

Дифференциальное исчисление имеет многочисленные приложения к исследованию изменения функций. Эти приложения будут рассмотрены в дальнейших параграфах этой главы.

§ 4. Правило Лопиталя

13. Неопределенности вида $\frac{0}{0}$. В п. III.7 мы говорили, что при вычислении предела отношения двух бесконечно малых могут получаться различные результаты. В первом печатном учебнике по дифференциальному исчислению (1696), написанном французским математиком Г. Лопита́лем (1661—1704), было опубликовано найденное И. Бернулли простое правило для вычисления такого предела, пригодное во многих случаях. Пусть ищется предел

$$\lim_{t \to t_0} \frac{\varphi(t)}{\psi(t)}, \tag{40}$$

причем

$$\varphi(t_0) = \psi(t_0) = 0, \tag{41}$$

т. е. мы имеем дело с неопределенностью вида $\frac{0}{0}$. Пусть каким-то способом найден предел

$$\lim_{t \to t_0} \frac{\varphi'(t)}{\psi'(t)} = k, \tag{42}$$

конечный или бесконечный. Тогда утверждается, что и предел (40) равен k, т. е. *для неопределенностей вида* $\frac{0}{0}$

$$\lim_{t \to t_0} \frac{\varphi(t)}{\psi(t)} = \lim_{t \to t_0} \frac{\varphi'(t)}{\psi'(t)} \tag{43}$$

— предел отношения функций равен пределу отношения производных.

Для доказательства рассмотрим линию $y=\varphi(t)$, $x=\psi(t)$ на плоскости x, y; тогда при $t \to t_0$ согласно (41) эта линия приближается к началу координат. Чтобы узнать, как именно она приближается (наподобие спирали или с определенного направления и с какого именно), заметим, что согласно (42)

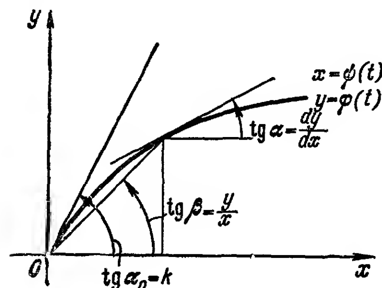


Рис. 119.

$$\frac{dy}{dx} = \frac{\varphi'(t)\,dt}{\psi'(t)\,dt} = \frac{\varphi'(t)}{\psi'(t)} \to k \qquad (\text{при } t \to t_0).$$

Значит (рис. 119), при приближении к началу координат касательная, поворачиваясь, в пределе становится под таким углом α_0 к оси x, что $\operatorname{tg}\alpha_0 = k$. Но тогда и «угол возвышения» β (см. рис. 119) при приближении к началу координат стремится к α_0, т. е.

$$\frac{\varphi(t)}{\psi(t)} = \frac{y}{x} = \operatorname{tg}\beta \underset{t \to t_0}{\longrightarrow} \operatorname{tg}\alpha_0 = k, \tag{44}$$

что и требовалось доказать.

Иногда при применении правила Лопиталя оказывается, что отношение производных снова является неопределенностью вида $\frac{0}{0}$; тогда это правило может быть вновь применено и т. д.

Например,

$$\lim_{x \to 0} \frac{x - \sin x}{x^3} = \left(\frac{0}{0}\right) = \lim_{x \to 0} \frac{1 - \cos x}{3x^2} = \left(\frac{0}{0}\right) =$$
$$= \lim_{x \to 0} \frac{\sin x}{6x} = \left(\frac{0}{0}\right) = \lim_{x \to 0} \frac{\cos x}{6} = \frac{1}{6};$$

$$\lim \frac{2^x - 4 \cdot 2^{-x}}{(x-1)^2} = \left(\frac{0}{0}\right) = \lim_{x \to 1} \frac{2^x \ln 2 + 4 \cdot 2^{-x} \ln 2}{2(x-1)} = \frac{4 \ln 2}{0} = \pm\infty,$$

в первом примере правило Лопиталя было применено три раза, а во втором — один раз.

Правило Лопиталя для неопределенности вида (40) всегда приводит к цели, если t_0 — конечное число, а числитель или знаменатель имеют *целый* порядок малости по сравнению с $t-t_0$ (см. п. III.10). Действительно, из того же правила Лопиталя можно вывести, что при каждом дифференцировании порядок малости понижается на единицу, а потому после нескольких шагов мы получим в числителе или знаменателе «нулевой порядок малости», т. е. конечный предел, не равный нулю, и неопределенности не будет.

14. Неопределенности вида $\frac{\infty}{\infty}$. Правило Лопиталя (43) сохраняется и для неопределенностей вида $\frac{\infty}{\infty}$, т. е. когда взамен условия (41) ставится условие

$$|\varphi(t_0)| = |\psi(t_0)| = \infty.$$

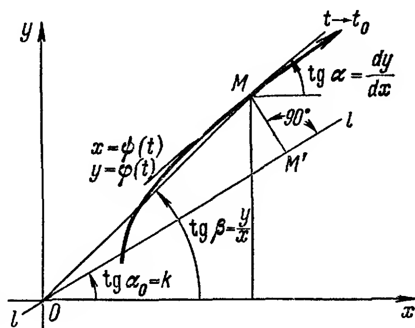


Рис. 120.

Доказательство аналогично проведенному в п. 13, однако теперь при $t \to t_0$ линия $y=\varphi(t)$, $x=\psi(t)$ не приближается к началу координат, а уходит в бесконечность (рис 120). При этом согласно условию (42) эта линия поворачивается так, что угол α, который она (т. е. ее касательная) составляет с осью x, стремится к α_0, где $\operatorname{tg}\alpha_0 = k$. Но тогда путь, пройденный точкой M этой линии вдоль прямой ll (см. рис. 120), будет бесконечно большой величиной более высокого порядка, чем путь поперек этого направления, точнее говоря, $MM' \ll OM$. Значит, при удалении точки M на бесконечность $\angle MOM' \to 0$ и поэтому «угол возвышения» β стремится к α_0 и мы можем написать те же формулы (44), чем доказательство и завершается.

Приведем несколько важных примеров:

$$\lim_{x\to\infty} \frac{\ln x}{x^k} = \left(\frac{\infty}{\infty}\right) = \lim_{x\to\infty} \frac{\frac{1}{x}}{kx^{k-1}} = \lim_{x\to\infty} \frac{1}{kx^k} = \frac{1}{\infty} = 0 \qquad (k>0);$$

$$\lim_{x\to\infty} \frac{x}{a^x} = \left(\frac{\infty}{\infty}\right) = \lim_{x\to\infty} \frac{1}{a^x \ln a} = \frac{1}{\infty} = 0 \qquad (a>1);$$

$$\lim_{x\to\infty} \frac{x^k}{b^x} = \lim_{x\to\infty} \left(\frac{x}{(\sqrt[k]{b})^x}\right)^k = \lim_{x\to\infty} \left(\frac{x}{a^x}\right)^k = 0^k = 0 \qquad (k>0, \quad b>1).$$

Таким образом, при стремлении аргумента к бесконечности *логарифмическая функция стремится к бесконечности медленнее, чем любая степенна́я* (с положительным показателем степени), *а степенна́я — медленнее, чем любая показательная* (с основанием, бо́льшим единицы).

К неопределенностям другого вида (см. п. III.5, свойство 6 и п. III.15.1) правило Лопиталя применяется после преобразования

их к виду $\frac{0}{0}$ или $\frac{\infty}{\infty}$. Это можно делать по схеме

$$0\cdot\infty=0\frac{1}{0}=\frac{0}{0},\qquad \infty-\infty=\frac{1}{0}-\frac{1}{0}=\frac{0-0}{0\cdot 0}=\frac{0}{0}$$

(это, конечно, условная запись, говорящая только о виде рассматриваемых величин). К степенны́м неопределенностям можно применять правило Лопиталя после их логарифмирования.

§ 5. Формула и ряд Тейлора

15. Формула Тейлора. В п. 10 мы видели, что замена приращения функции ее дифференциалом дает возможность получить многие приближенные формулы. Оказывается, что эти формулы можно значительно уточнить, если применить дифференциалы высшего порядка; об этом и будет говорить формула Тейлора, названная по имени английского математика Б. Те́йлора (1685—1731).

Предположим сначала, что нам дан многочлен $P(x)$. Обычно он считается разложенным по степеням x, но его без труда можно разложить и по степеням $x-a$, где a — какое угодно число.

Пусть, например, мы хотим разложить многочлен $P(x)=5-3x+2x^3$ по степеням $x-4$. Для этого надо в $P(x)$ подставить $x=[4+(x-4)]$ и затем раскрыть квадратные скобки, не раскрывая круглых:

$$P(x)=5-3[4+(x-4)]+2[4+(x-4)]^3=$$
$$=5-12-3(x-4)+128+96(x-4)+24(x-4)^2+2(x-4)^3=$$
$$=121+93(x-4)+24(x-4)^2+2(x-4)^3.$$

В общем случае для многочлена степени n можно написать

$$P(x)=a_0+a_1(x-a)+a_2(x-a)^2+a_3(x-a)^3+\ldots+a_n(x-a)^n. \tag{45}$$

Коэффициенты здесь можно найти так. Положив сначала $x=a$, получим $P(a)=a_0$. Продифференцируем затем формулу (45):

$$P'(x)=a_1+2a_2(x-a)+3a_3(x-a)^2+\ldots+na_n(x-a)^{n-1}.$$

Если здесь положить $x=a$, получим $P'(a)=a_1$. Продифференцируем еще раз:

$$P''(x)=1\cdot 2a_2+2\cdot 3a_3(x-a)+\ldots+(n-1)na_n(x-a)^{n-2};$$

отсюда $P''(a)=1\cdot 2a_2$. Далее, аналогично получим $P'''(a)=1\cdot 2\cdot 3a_3$ и т. д., в общем случае $P^{(k)}(a)=k!a_k$ (где $k!=1\cdot 2\ldots k$), откуда

$$a_k=P^{(k)}(a)/k!$$

Итак, формулу (45) можно переписать в таком виде:

$$P(x)=$$
$$=P(a)+\frac{P'(a)}{1!}(x-a)+\frac{P''(a)}{2!}(x-a)^2+\ldots+\frac{P^{(n)}(a)}{n!}(x-a)^n=$$
$$=P(a)+\sum_{k=1}^{n}\frac{P^{(k)}(a)}{k!}(x-a)^k, \quad (46)$$

где $\sum$ — знак суммирования (см. п. III.6).

Так, в приведенном примере

$$P'(x)=-3+6x^2, \quad P''(x)=12x, \quad P'''(x)=12,$$
$$P(4)=5-3\cdot 4+2\cdot 4^3=121, \quad \frac{P'(4)}{1!}=-3+6\cdot 4^2=93,$$
$$\frac{P''(4)}{2!}=\frac{12\cdot 4}{2}=24, \quad \frac{P'''(4)}{3!}=\frac{12}{6}=2,$$

т. е. получаем те же значения коэффициентов, что и выше.

Пример. Положим $P(x)=(s+x)^n$, $a=0$ и применим формулу (46). Тогда

$$P'(x)=n(s+x)^{n-1};$$
$$P''(x)=n(n-1)(s+x)^{n-2};$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$P^{(n)}(x)=n(n-1)\ldots 1$$

и потому по формуле (46)

$$(s+x)^n=s^n+\frac{n}{1!}s^{n-1}x+\frac{n(n-1)}{2!}s^{n-2}x^2+\ldots+\frac{n(n-1)\ldots 2\cdot 1}{n!}x^n.$$

Таким образом, мы получаем явное выражение для биномиальных коэффициентов, о которых мы упоминали в п. 11.

Если теперь вместо многочлена $P(x)$ взять произвольную функцию $f(x)$, то формула (46) уже не будет справедлива; но если обозначить отличие левой части формулы (46) от правой через $R_n(x)$ (*остаточный член*), то можно написать

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots$$
$$\ldots+\frac{f^{(n)}(a)}{n!}(x-a)^n+R_n(x). \quad (47)$$

Это и есть *формула Тейлора*. Существенно, что в ней *при $x\to a$ остаточный член имеет по крайней мере $(n+1)$-й порядок малости по сравнению с $x-a$*, т. е. более высокий порядок, чем последний из выписанных «точных» членов в формуле (47). Для доказательства этого положим для простоты, что, например, $n=2$, т. е.

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+R_2(x).$$

Выражая отсюда $R_2(x)$ и применяя правило Лопиталя (п. 13), получим

$$\lim_{x\to a}\frac{R_2(x)}{(x-a)^3}=\lim_{x\to a}\frac{f(x)-f(a)-f'(a)(x-a)-\frac{f''(a)}{2!}(x-a)^2}{(x-a)^3}=$$

$$=\left(\frac{0}{0}\right)=\lim_{x\to a}\frac{f'(x)-f'(a)-f''(a)(x-a)}{3(x-a)^2}=\left(\frac{0}{0}\right)=$$

$$=\lim_{x\to a}\frac{f''(x)-f''(a)}{3\cdot 2(x-a)}=\left(\frac{0}{0}\right)=\lim_{x\to a}\frac{f'''(x)}{3!}=\frac{f'''(a)}{3!} \qquad (48)$$

(конечный предел).

Отсюда (см. п. III.10) и вытекает наше утверждение о порядке малости $R_2(x)$, а аналогично — и $R_n(x)$.

Если обозначить $x=a+h$, то, обрывая формулу (47) все дальше и дальше, получим все более точные (при малых $|h|$) формулы:

$$f(a+h)\approx f(a)+f'(a)h \qquad (49)$$

с точностью до величины порядка h^2 (это — формула (27));

$$f(a+h)\approx f(a)+f'(a)h+\frac{f''(a)}{2}h^2 \qquad (50)$$

с точностью до величины порядка h^3,

$$f(a+h)\approx f(a)+f'(a)h+\frac{f''(a)}{2}h^2+\frac{f'''(a)}{6}h^3 \qquad (51)$$

с точностью до величины порядка h^4 и т. д.

Многочлены (относительно $h=x-a$), стоящие в правых частях, называются *многочленами Тейлора*. Они дают в некотором смысле наилучшее приближенное выражение функции $f(x)$ в виде многочлена данной степени вблизи значения $x=a$. Именно, среди всех многочленов этой степени многочлен Тейлора отличается от $f(x)$ на величину наивысшего порядка малости при $x\to a$. Например, если в правой части формулы (50) изменить хотя бы один коэффициент, то отличие будет уже в величинах 0, 1 или 2-го порядка малости, а не 3-го, как для многочлена Тейлора.

16. Ряд Тейлора. Так как ошибки формул (49), (50), (51) и т. д. становятся все более высокого порядка малости, то естественно предположить, что при малых $|h|$ можно перейти к пределу и получить *точное* представление $f(a+h)$ в виде суммы бесконечного ряда (см. п. III.6)

$$f(a+h)=f(a)+\frac{f'(a)}{1!}h+\frac{f''(a)}{2!}h^2+\ldots+\frac{f^{(n)}(a)}{n!}h^n+\ldots$$

$$\ldots=f(a)+\sum_{n=1}^{\infty}\frac{f^{(n)}(a)}{n!}h^n. \qquad (52)$$

Этот ряд называется *рядом Тейлора*; он был впервые получен Б. Тейлором в 1715 г. В § 3 гл. XVII, где будут систематически

изучаться такие ряды, мы увидим, что это предположение оправдывается; в частности, тогда будет выяснен вопрос, для каких именно h в принципе можно пользоваться формулой (52). Оказывается, что это можно делать всегда, если ряд «практически сходится» в том смысле, как это описано в конце п. III.6 (однако при этом функция, разлагаемая в ряд Тейлора, не должна задаваться различными формулами на разных участках изменения аргумента, см. п. I.13). На основании этого мы будем применять ряды Тейлора уже сейчас.

Формулу (52) можно переписать в виде ($a+h=x$, $h=x-a$):

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots \tag{53}$$

(*разложение по степеням* $x-a$). В частности, если $a=0$, получим *разложение по степеням* x

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\ldots, \tag{54}$$

которое иногда исторически неправильно называется «*рядом Маклорена*» по имени шотландского математика К. Маклóрена (1698—1746).

Пусть, например $f(x)=e^x$. Тогда $f'(x)=e^x$, $f''(x)=e^x, \ldots,$

$$f(0)=1,\quad f'(0)=1,\quad f''(0)=1, \ldots$$

и применение формулы (54) дает

$$e^x=1+\frac{1}{1!}x+\frac{1}{2!}x^2+\frac{1}{3!}x^3+\ldots \tag{55}$$

Вычислим число e с точностью до 0,001. Для этого положим $x=1$ и подсчитаем члены один за другим с одним лишним знаком:

$$e=1,0000+1,0000+0,5000+0,1667+0,0417+0,0083+\\+0,0014+0,0002+0,0000.$$

Здесь каждый следующий член получается из предыдущего делением на очередное целое число. Как видим, члены ряда обнаруживают явную тенденцию к быстрому убыванию, причем выходят за рамки принятой степени точности. Складывая и округляя, получим $e=2,718$.

Аналогично (55) можно получить формулы, вывести которые мы предоставим читателю:

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots; \tag{56}$$

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\ldots; \tag{57}$$

$$\operatorname{ch} x=1+\frac{x^2}{2!}+\frac{x^4}{4!}+\frac{x^6}{6!}+\ldots; \tag{58}$$

$$\operatorname{sh} x=x+\frac{x^3}{3!}+\frac{x^5}{5!}+\frac{x^7}{7!}+\ldots; \tag{59}$$

$$(1+x)^a=1+\binom{a}{1}x+\binom{a}{2}x^2+\binom{a}{3}x^3+\ldots \tag{60}$$

(при любом a), где *биномиальные коэффициенты* $\binom{a}{k}$ определяются по формуле $\binom{a}{1}=a$, $\binom{a}{2}=\frac{a(a-1)}{1\cdot 2}, \ldots, \binom{a}{k}=$ $=\frac{a(a-1)\ldots(a-k+1)}{k!}$;

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4!}+\ldots \qquad (61)$$

Ряд Тейлора можно переписать в другом виде, если обозначить $x-a=\Delta x$; $f(x)-f(a)=\Delta f$; $f'(a)(x-a)=f'(a)\Delta x=df$; $f''(a)(x-a)^2=f''(a)(\Delta x)^2=d^2f$ (см. п. 12) и т. д. Получим из (53)

$$\Delta f=df+\frac{d^2f}{2!}+\frac{d^3f}{3!}+\ldots+\frac{d^nf}{n!}+\ldots \qquad (62)$$

Обрывая эту формулу, получим все более и более точные (при малых Δx) формулы: $\Delta f\approx df$ с точностью до величины порядка $(\Delta x)^2$; $\Delta f\approx df+\frac{1}{2}d^2f$ с точностью до величины порядка $(\Delta x)^3$ и т. д.

§ 6. Интервалы монотонности и экстремум

17. Знак производной. Пусть рассматривается функция $y=f(x)$; *в этом пункте мы будем предполагать, что как она сама, так и ее производная не имеют разрывов.* Примерный график этой функции показан на рис. 121. Так как $y'=\operatorname{tg}\alpha$ (см. п. 3), то *функция возрастает в каждом интервале, в котором ее производная положительна, и убывает в каждом интервале, в котором ее производная отрицательна.* Другими словами, если скорость изменения какой-либо величины положительна, то эта величина возрастает, а если скорость отрицательна, то величина убывает.

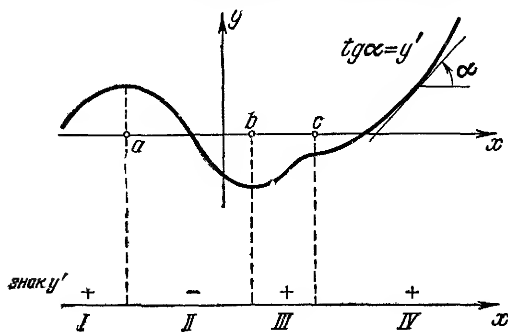


Рис. 121.

Так как производная, переходя непрерывно от положительных значений к отрицательным, должна пройти через нулевое значение, то в точках, в которых интервал возрастания сменяется интервалом убывания, будет $y'=0$. Точки x, в которых $f'(x)=0$, называются *стационарными точками* функции f: в них мгновенная скорость изменения функции равна нулю, т. е. это как бы точки мгновенного

покоя. На рис. 121 имеются три стационарные точки: a, b и c. Соответствующие значения функции называются ее *стационарными значениями*.

Из сказанного следует, что для нахождения интервалов монотонности функции $f(x)$ надо на ось x нанести все стационарные точки этой функции, после чего проверить знак f' на каждом из интервалов между соседними стационарными точками. Интервалы, на которых $f'>0$, будут интервалами возрастания, а интервалы, на которых $f'<0$, будут интервалами убывания функции f. При этом если на двух соседних интервалах знак f' одинаков, то они составляют единый интервал монотонности; так, на рис. 121 интервалы *III* и *IV* составляют единый интервал возрастания функции $f(x)$.

Очевидно также, что *интервалами постоянства функции f служат те и только те интервалы, на которых* $f'(x)\equiv 0$, так как на этих интервалах функция f не может ни возрастать, ни убывать.

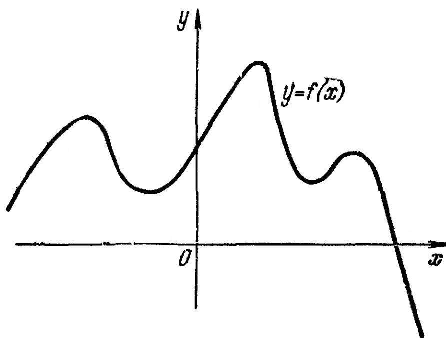


Рис. 122.

18. Точки экстремума. *Если при некотором $x=x_0$ значение $f(x_0)$ больше всех «соседних» значений функции* (т. е. значений $f(x)$ при x, достаточно близких к x_0), *то точка $x=x_0$ называется точкой максимума функции f, а значение $f(x_0)$ — ее максимальным значением.* Аналогично определяются *точка минимума* и *минимальное значение* функции. Так, на рис. 121 функция имеет точку максимума при $x=a$ и точку минимума при $x=b$. В других примерах функция может иметь другое количество точек максимума и минимума, причем у непрерывной функции они обязательно чередуются. Так, на рис. 122 функция имеет три точки максимума и две — минимума; на рис. 46 тех и других бесконечное число, а на рис. 44 их совсем нет.

Как максимум, так и минимум объединяются словом «*экстре́мум*» от латинского extremus, что значит «крайний». Из п. 17 следует, что *точками экстремума служат точки, при переходе через которые производная сменяет знак*. Более определенно: *если при переходе x через точку $x=a$ в положительном направлении знак $f'(x)$ сменяется с + на —, то при $x=a$ функция f имеет максимум*, так как при этом переходе функция f сменяет возрастание на убывание (см. рис. 121). Аналогично *при переходе через точку минимума производная сменяет знак с — на +*.

Из п. 17 следует также, что в указанных там предположениях *все точки экстремума функции будут ее стационарными точками*. Этот

необходимый признак экстремума получил, по существу, уже Ферма. Как видно из рис. 121, признак не является достаточным, т. е. стационарная точка может и не быть точкой экстремума.

Достаточными признаками пользуются реже, чем необходимым, так как во многих конкретных задачах часто бывает заранее ясно, что экстремум должен быть, и даже примерно, где он будет, только точное его значение неизвестно. Если при этом необходимый признак даст лишь одну возможную точку, то, значит, там и будет экстремум. Если экстремумов несколько, то их можно находить одновременно с отысканием интервалов монотонности, как об этом говорилось в п. 17.

Так как вблизи стационарной точки значения функции меняются вёсьма медленно, то из признака Ферма вытекает, что если точка экстремума найдена с некоторой погрешностью, то погрешность в соответствующем экстремальном значении будет высшего порядка малости. Поэтому выгодно, если только это возможно, приводить задачу о нахождении той или иной величины к задаче о нахождении экстремального или просто стационарного значения некоторой функции. Тогда даже грубое отыскание точки экстремума даст хороший окончательный результат.

Условия экстремума можно получить также с помощью формулы Тейлора (см. п. 15). Будем исследовать точку $x=a$ для функции $f(x)$. Тогда из формулы (49) видим, что если $f'(a)\neq 0$, то экстремума при $x=a$ нет, так как, меняя знак у h, мы изменим знак и у $f'(a)h$, т. е. у разности $f(a+h)-f(a)$ (ибо величины порядка h^2 при малых h ничтожны по сравнению с величиной $f'(a)h$).

Если $f'(a)=0$, $f''(a)\neq 0$, то из формулы (50) аналогичным образом находим, *что экстремум при* $x=a$ *будет: минимум, если* $f''(a)>0$ (тогда $f(a+h)>f(a)$ при малых $|h|$), *и максимум, если* $f''(a)<0$. Если $f'(a)=0$, $f''(a)=0$, $f'''(a)\neq 0$, то из формулы (51) следует, что экстремума при $x=a$ не будет; если $f'(a)=0$, $f''(a)=0$, $f'''(a)=0$, $f^{IV}(a)\neq 0$, то экстремум будет и т. д.

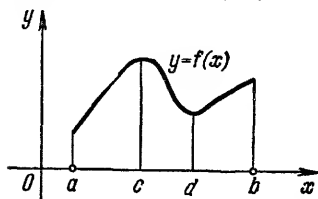


Рис. 123.

19. Наибольшее и наименьшее значения функции. Пусть, как и выше, функция $f(x)$ непрерывна вместе со своей производной на некотором замкнутом интервале $a\leqslant x\leqslant b$ и требуется найти на нем *наибольшее и наименьшее значения этой функции*. Тогда наряду с рассмотренными в п. 18 *внутренними* экстремумами надо принять во внимание также *краевые* экстремумы: так, на рис. 123 функция имеет краевой минимум при $x=a$ и краевой максимум при $x=b$ наряду с двумя внутренними экстремумами. Конечно, в точках краевого экстремума производная не обязана равняться нулю.

Далее, надо иметь в виду, что в п. 18 рассматривались *локáльные* (местные, от латинского locus — место) экстремумы, а здесь нас интересуют *тотáльные* (от латинского totalis — всеобъемлющий) максимум

и минимум. Поэтому для нахождения наибольшего значения функции на интервале $a \leqslant x \leqslant b$ надо найти все ее внутренние и концевые максимумы на этом интервале, а затем сравнить между собой все соответствующие максимальные значения: наибольший из максимумов и даст наибольшее значение функции. Аналогично находится наименьшее значение функции на замкнутом интервале. Для упрощения работы можно просто сравнить все стационарные и краевые значения функции: наибольшее из них даст тотальный максимум, а наименьшее — тотальный минимум.

Если функция $f(x)$ непрерывная, но f' может иметь разрывы, то смена возрастания f на ее убывание может происходить не только в точках, где $f'=0$, но и в точках, где f' имеет разрыв. Для нахождения интервалов монотонности функции f, т. е. для выяснения знака f', надо поступать аналогично тому, как в п. III.15 выяснялся знак f. Если при переходе через некоторое значение $x=a$ производная $f'(x)$ меняет знак, претерпевая разрыв, то при значении $x=a$ функция $f(x)$ имеет *острый экстремум* (см. острый минимум на рис. 35 и острые максимумы на рис. 106). Вблизи острого экстремума функция f, конечно, уже не будет медленно меняющейся, как вблизи стационарных экстремумов, рассмотренных в п. 18.

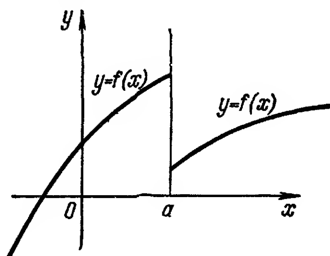


Рис. 124.

Итак, полная формулировка необходимого условия экстремума для непрерывной функции такова: *в точке экстремума производная обращается в нуль или терпит разрыв.*

Естественно, что для точек острого экстремума условия, основанные на применении формулы Тейлора, отпадают; остается условие, основанное на перемене знака производной.

Если сама функция $f(x)$ имеет разрывы, то точки разрыва могут служить концами интервалов монотонности этой функции, даже если по обе стороны от точки разрыва f' имеет одинаковый знак. Так, на рис. 124 всюду при $x \neq a$ будет $y' > 0$ и в то же время имеются два интервала возрастания f: $-\infty < x < a$ и $a < x < \infty$, которые нельзя объединить в один. Поэтому при нахождении интервалов монотонности на ось x надо нанести также все точки разрыва функции.

При нахождении наибольшего значения функции, имеющей разрывы, надо иметь в виду, что такая функция может получиться неограниченной сверху и тогда наибольшего значения, конечно, не будет. То же осложнение может возникнуть при рассмотрении функции, даже непрерывной, но на бесконечном интервале.

Но даже если такая ограниченность и будет, то о достижении наибольшего значения при наличии точек разрыва или в случае неограниченности интервала часто можно говорить только в предельном смысле. Так, на рис. 124 наибольшим значением является $f(a-0)$, причем малейший переход через точку a приводит к резкому уменьшению этого значения, которое, таким образом, является «неустойчивым». В этом случае предпочитают говорить не о наибольшем значении, а о *«верхней границе»* значений функции, понимая под этим термином наибольшее из всех значений функции и из всех ее пределов.

Пример 1. Пусть функция $y=f(x)=(1+x^2)/(1+x^4)$ рассматривается на всей оси x. Точек разрыва ни она, ни ее производная

$$y'=\frac{2x(1+x^4)-4x^3(1+x^2)}{(1+x^4)^2}=2x\frac{1-2x^2-x^4}{(1+x^4)^2}$$

не имеют, и поэтому для нахождения интервалов монотонности надо приравнять $y'=0$, что даст уравнение $x(1-2x^2-x^4)=0$, т. е.

$$x_1=0;\quad x^4+2x^2-1=0;\quad (x^2)^2+2x^2-1=0;\quad x^2=-1\pm\sqrt{2};$$

годится только $+$, т. е.

$$x^2=\sqrt{2}-1;\quad x_{2,3}=\pm\sqrt{\sqrt{2}-1}=\pm 0{,}644$$

(рис. 125). Таким образом, ось x разбивается на четыре интервала. Подстановка в y' значений $x=-10$; $-0{,}1$; $0{,}1$; 10 из этих интервалов дает соответственно знаки $+$, $-$, $+$, $-$. Значит, эти интервалы последовательно являются интервалами возрастания, убывания, возрастания и убывания. При $x=x_3$, x_2 функция имеет максимумы, а при $x=x_1$—минимум. Максимальные значения

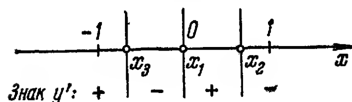


Рис. 125.

$$f(x_2)=f(x_3)=\frac{1+(\sqrt{2}-1)}{1+(\sqrt{2}-1)^2}=$$
$$=\frac{\sqrt{2}}{4-2\sqrt{2}}=\frac{\sqrt{2}+1}{2}=1{,}207,$$

а минимальное значение $f(x_1)=1$.

Кроме того, «концевые» предельные значения $f(\infty)=f(+\infty)=0$, так как при $x\longrightarrow\pm\infty$ в числителе у $f(x)$ получается бесконечно большая величина второго порядка в сравнении с x, а в знаменателе—четвертого. Значит, наибольшее значение 1,207 функции достигается при $x=\pm 0{,}644$, а наименьшее значение нуль—только в пределе при $x\longrightarrow\pm\infty$. Примерный график функции показан на рис. 126.

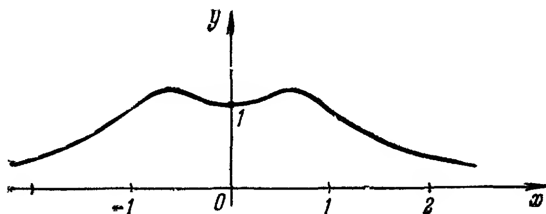


Рис. 126.

Пример 2. Пусть из квадратного листа жести со стороной a требуется выкроить коробку наибольшей вместимости. (См. рис. 127; линии сгиба проведены пунктиром, а линии разреза—сплошные.) Ясно, что какое-то решение этой задачи имеется, но неясно, где проводить разрез (т. е. каково x) и какой получится объем. Если сначала принять x каким-то неопределенным, то объем $V=(a-2x)^2x$, причем по смыслу задачи x должен быть между 0 и $\frac{a}{2}$. Применение необходимого признака экстремума дает

$$\frac{dV}{dx}=2(a-2x)(-2)x+(a-2x)^2 1=(a-2x)(a-6x)=0,$$

откуда $x_1=a/2$, $x_2=a/6$. По смыслу задачи подходит только $x=a/6$, т. е. там и будет максимум. Максимальный объем

$$V_{\max}=\left(a-2\cdot\frac{a}{6}\right)^2\frac{a}{6}=\frac{2}{27}a^3.$$

Пример 3. Рассмотрим задачу о *преломлении света* на границе раздела двух *однородных* (т. е. одинаковых во всех своих точках) *изотропных* (т. е. одинаковых во всех направлениях) сред. Предположим сначала, что граница раздела плоская; проведем через луч света плоскость (рис. 128) и

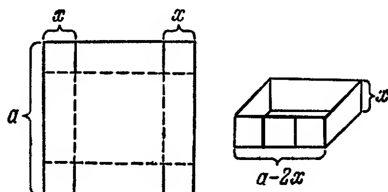


Рис. 127.

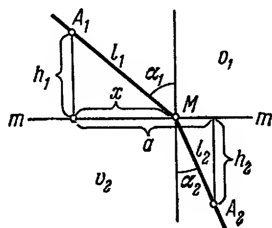


Рис. 128.

v_1 — скорость света в первой среде; v_2 — скорость света во второй среде; mm — граница раздела сред.

выберем на луче точки A_1 и A_2. Воспользуемся, далее, *принципом Ферма́ в оптике*, который гласит: из всех возможных путей, идущих из A_1 в A_2, луч света выбирает такой, который он проходит за минимальное время. Поэтому точка M при заданных A_1 и A_2 должна быть расположена так, чтобы

$$t=t_1+t_2=\frac{l_1}{v_1}+\frac{l_2}{v_2}=\frac{\sqrt{h_1^2+x^2}}{v_1}+\frac{\sqrt{h_2^2+(a-x)^2}}{v_2}$$

было минимально возможным. Применяя необходимое условие минимума, получим

$$\frac{dt}{dx}=\frac{x}{v_1\sqrt{h_1^2+x^2}}-\frac{a-x}{v_2\sqrt{h_2^2+(a-x)^2}}=0,$$

откуда

$$\frac{x}{l_1v_1}=\frac{a-x}{l_2v_2},\quad \frac{\sin\alpha_1}{\sin\alpha_2}=\frac{x}{l_1}:\frac{a-x}{l_2}=\frac{v_1}{v_2}.$$

Итак, получаем закон преломления: отношение синуса угла падения к синусу угла преломления есть величина постоянная, равная отношению скоростей света в обеих средах. Если теперь поверхность раздела не плоская, то так как закон преломления зависит лишь от ситуации в бесконечной близости от точки преломления, а в такой близости поверхность раздела можно считать плоской, то и в этом случае закон преломления остается тем же.

Таким образом, мы видим, *что закон физики удалось с помощью решения задачи на экстремум вывести из общего физического принципа, имеющего экстремальный характер, т. е. утверждающего экстремальное значение определенной величины в реальных условиях.*

Более подробное исследование показывает, что в принципе Ферма, как и в ряде других аналогичных принципов, существенна не минимальность и даже не экстремальность времени прохождения светом пути, а стационарность этого времени. В такой форме этот принцип можно вывести из волновой теории света.

§ 7. Построение графиков

Дифференциальное исчисление дает общий метод выявления индивидуальных особенностей графика заданной функции $y=f(x)$, что позволяет строить этот график быстрее и точнее, чем «по точкам», как это делалось в п. I.14. Так, нахождение интервалов монотонности функции и точек ее экстремума, описанное в § 6, существенно при этом построении. Кроме этого, полезными оказываются еще некоторые исследования; о них мы сейчас будем говорить.

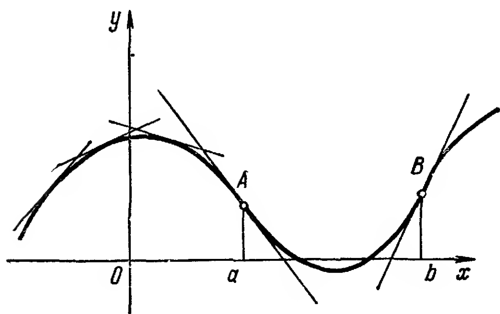


Рис. 129.

20. Участки выпуклости графика и точки перегиба. Пусть график $y=f(x)$ таков, как показано на рис. 129. Мы видим, что левее точки A и правее точки B график выпуклый кверху, а между A и B — выпуклый книзу (см. п. I.24). Точки A и B, в которых выпуклость в одну сторону сменяется выпуклостью в другую сторону, являются точками перегиба; в них график пересекает касательную, хотя и под нулевым углом.

Для нахождения участков выпуклости кверху и книзу заметим, что на участке выпуклости кверху (например, на рис. 129 при $x<a$) с ростом x касательная к графику поворачивается по часовой стрелке, т. е. ее угловой коэффициент убывает. Но этот коэффициент равен y' и, таким образом, график будет выпуклым кверху и книзу для тех интервалов оси x, для которых y' соответственно убывает или возрастает. Эти интервалы находятся с помощью исследования знака y'' в точности так же, как в п. 17 интервалы убывания и возрастания y находились с помощью исследования знака y'. Итак, *график будет выпуклым кверху и книзу для тех интервалов оси x, для которых соответственно $y''<0$ и $y''>0$; точки перегиба получаются при тех значениях x, при переходе через которые y'' меняет знак. В самой же точке перегиба производная y'' равна нулю.* При этом предполагается, что y, y' и y'' не имеют разрывов. Если такие разрывы имеются, то интервалы выпуклости кверху и книзу графика строятся после нанесения на ось x всех разрывов, поскольку они также могут служить концами названных интервалов.

21. Асимптоты графика. Асимптоты графика $y=f(x)$ могут быть вертикальные (параллельные оси y) и невертикальные (рис. 130). Первых может быть сколько угодно, даже бесконечное число (тан-

генсоида, рис. 47), и они находятся так: если $|y| \to \infty$ при $x \to a$ (a конечное), то прямая $x = a$ служит вертикальной асимптотой.

Невертикальных асимптот не может быть более двух (одной при $x \to \infty$ и одной при $x \to -\infty$), и они находятся так: пусть прямая $y = kx + b$ служит асимптотой графика $y = f(x)$ при $x \to \infty$. Тогда (см. рис. 130) разность $\delta = y_{\text{асимптоты}} - y_{\text{графика}}$ равна

$$\delta = (kx + b) - f(x) \qquad (63)$$

и стремится к нулю при $x \to \infty$, откуда

$$\frac{f(x)}{x} = k + \frac{b}{x} - \frac{\delta}{x} \to k,$$

т. е.

$$k = \lim_{x \to \infty} \frac{f(x)}{x}.$$

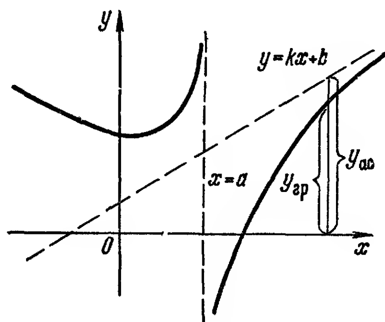


Рис. 130.

Кроме того, в силу (63)

$$f(x) - kx = b - \delta \to b,$$

т. е.

$$b = \lim_{x \to \infty} [f(x) - kx].$$

Каждый из этих двух пределов должен существовать и быть конечным, в противном случае асимптоты при $x \to \infty$ нет. *Если же эти конечные пределы существуют, то и асимптота существует*, так как из последнего равенства видно, что величина

$$[f(x) - kx] - b \to 0, \text{ т. е. } f(x) - (kx + b) \to 0 \quad \text{при} \quad x \to \infty.$$

22. Общая схема исследования функции и построения ее графика. Эта схема для функции $y = f(x)$ включает следующее:

1) Ищутся область определения функции, точки разрыва и нули, после чего выясняются интервалы положительности и отрицательности функции. Выясняется поведение функции при приближении к ее точкам разрыва и к концам интервалов, на которых функция определена (в том числе поведение функции на бесконечности). Находятся асимптоты графика. Выясняется, не будет ли функция четной, нечетной, периодической и т. п.

2) Ищутся точки разрыва и нули производной, после чего выясняются интервалы возрастания и убывания функции, точки экстремума и экстремальные значения. Выясняется поведение производной при приближении к ее точкам разрыва, к точкам разрыва функции (если функция в них имеет конечный скачок) и к краям интервалов, на которых функция определена (если эти края конечные и функция там имеет конечное значение).

3) Ищутся точки разрыва и нули второй производной, после чего выясняются участки выпуклости кверху и книзу графика, а также точки перегиба, в которых полезно найти направление касательной.

Все найденные точки наносятся на координатную плоскость, после чего строится сам график, в поведении которого должны быть переданы все найденные индивидуальные особенности. Если из них поведение графика недостаточно ясно, то надо построить еще несколько точек графика, вычислив значения y для отдельных значений x; желательно также, вычислив значения y', найти в этих точках направление касательной.

Приведем в качестве примера исследование графика функции $y=\sqrt[3]{x^3-2x^2}$. В данном случае областью определения служит вся ось $-\infty<x<\infty$; точек разрыва нет. Приравнивание $y=0$ показывает, что функция имеет два нуля, $x_1=0$ и $x_2=2$, т. е. получается три интервала знакопостоянства: $-\infty<x<0$, $0<x<2$, $2<x<\infty$. Подстановка произвольных значений из этих интервалов показывает, что на первом и втором функция отрицательна, а на третьем — положительна. Вертикальных асимптот нет. Вычисление невертикальных асимптот, проведенное в соответствии с п. 21, показывает (проверьте!), что одна и та же прямая $y=x-2/3$ является асимптотой графика как при $x\to\infty$, так и при $x\to-\infty$.

Вычислив производную $y'=\dfrac{3x^2-4x}{3\sqrt[3]{(x^3-2x^2)^2}}=\dfrac{3x-4}{3\sqrt[3]{(x-2)^2x}}$, видим, что она имеет разрывы (обращается в бесконечность) при $x=0$ и $x=2$ и равна нулю при $x=4/3$. Получается четыре интервала монотонности: $-\infty<x<0$; $0<x<\dfrac{4}{3}$; $\dfrac{4}{3}<x<2$; $2<x<\infty$, и подстановка в y' произвольных значений из этих интервалов показывает, что интервалом убывания служит только второй, а остальные служат интервалами возрастания; поэтому третий и четвертый образуют единый интервал возрастания. Таким образом, смена характера монотонности происходит при $x=0$ (максимум с максимальным значением $y=0$) и при $x=4/3$ (минимум с минимальным значением $y=-2\sqrt[3]{4}/3=-1,058$).

Вычислив вторую производную, получим после преобразований

$$y''=-\frac{8}{9\sqrt[3]{(x-2)^5x^4}}$$

(проверьте!). Она имеет разрыв там же, где первая производная (при $x=0$ и $x=2$), и совсем не имеет нулей. Получаются три участка «одинаковой выпуклости»: $-\infty<x<0$, $0<x<2$, $2<x<\infty$. Подстановка показывает, что на первом и втором участках выпуклость направлена книзу, а на третьем — кверху. Вычислим еще при $x=-1$:

$$y=-\sqrt[3]{3}=-1,44;\qquad y'=\frac{7}{3\sqrt[3]{9}}=1,12;$$

при $x=1$:

$$y=-1;\quad y'=-\frac{1}{3}=-0,33;$$

при $x=3$:

$$y=\sqrt[3]{9}=2,08;\qquad y'=\frac{5}{3\sqrt[3]{3}}=1,16.$$

Получающийся график изображен на рис. 131, причем вычисленные его точки показаны кружками (проверьте, что на графике переданы все его индивидуальные особенности).

Расположение графика относительно своей асимптоты легко вывести из *асимптотического разложения* рассматриваемой функции, т. е. разложения,

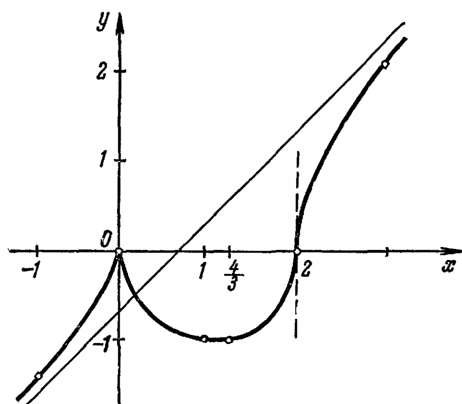


Рис. 131.

справедливого при достаточно больших $|x|$; это разложение в свою очередь вытекает из формулы (60):

$$y=\sqrt[3]{x^3-2x^2}=\sqrt[3]{x^3\left(1-\frac{2}{x}\right)}=x\left(1-\frac{2}{x}\right)^{1/3}=$$
$$=x\left[1+\frac{\frac{1}{3}}{1!}\left(-\frac{2}{x}\right)+\frac{\frac{1}{3}\left(\frac{1}{3}-1\right)}{2!}\left(-\frac{2}{x}\right)^2+\ldots\right]=$$
$$=x-\frac{2}{3}-\frac{4}{9x}+\text{члены высшего порядка малости.} \quad (64)$$

Значит, при больших $x>0$ будет $y<x-\frac{2}{3}=y_{\text{ас}}$, а при больших $x<0$ будет $y>y_{\text{ас}}$. Кроме того, из равенства (64) непосредственно следует, что

$$y-\left(x-\frac{2}{3}\right)\xrightarrow[|x|\to\infty]{}0. \quad (65)$$

Значит, если бы мы раньше не знали уравнения асимптоты $y=x-\frac{2}{3}$, мы вывели бы его из соотношения (65). Таким образом, мы получили еще один метод нахождения асимптоты.

ПРИБЛИЖЕННОЕ РЕШЕНИЕ КОНЕЧНЫХ УРАВНЕНИЙ. ИНТЕРПОЛЯЦИЯ

§ 1. Приближенное решение конечных уравнений

1. Введение. Мы будем рассматривать здесь численное решение уравнений вида

$$f(x)=0, \tag{1}$$

где f — заданная функция. Такие уравнения могут быть *алгебраическими*, если функция f алгебраическая, или *трансцендентными* в противном случае; как те, так и другие называются *конечными* в отличие, например, от дифференциальных уравнений. Мы укажем лишь несколько наиболее универсальных методов решения уравнений вида (1); другие методы читатель может найти в курсах приближенных вычислений.

Численное решение уравнения (1) обычно начинают с нахождения грубого, совсем приближенного решения, так называемого *нулевого приближения*. Если решается физическая задача, то это решение может быть известно из физического смысла задачи. Можно начать с примерного, хотя бы довольно грубого построения графика функции f. Если при этом обнаружится, что на каком-нибудь интервале a, b эта функция всюду определена, не имеет точек разрыва и принимает в точках a и b значения противоположных знаков, то в силу свойств непрерывных функций (п. III.14) f должна иметь на этом интервале хотя бы один нуль, т. е. уравнение (1) имеет там по крайней мере один корень. Если к тому же функция f на этом интервале монотонна, то такой корень здесь только один, т. е. этот корень *отделен* от остальных. Если обозначить этот, пока неизвестный, корень через α, то можно ручаться, что $a<\alpha<b$. Для дальнейшего уточнения значения α применяются различные методы (см. п. 2).

Часто бывает удобнее переписать уравнение (1) в форме $\varphi(x)=\psi(x)$, после чего искать пересечение графиков $y=\varphi(x)$ и $y=\psi(x)$. При этом левую часть уравнения (1) стараются разбить так, чтобы получились хорошо известные или, во всяком случае, более или менее простые графики; иногда помогает замена неизвестной x.

Рассмотрим, например, уравнение

$$\operatorname{tg} ax^2-bx^2=0, \tag{2}$$

где a, b — заданные положительные постоянные. Замена $ax^2 = s$ приводит к уравнению

$$\operatorname{tg} s = ks \qquad \left(k = \frac{b}{a}\right). \tag{3}$$

Графики левой и правой частей показаны на рис. 132, причем ясно, что нас интересуют лишь значения $s \geqslant 0$. Мы видим, что уравнение (3), а с ним и (2), имеет бесконечное число решений $s_0 = 0 < s_1 < s_2 < \ldots$, причем на графике хорошо видна зависимость решений от k, т. е. от a и b. В частности, видно, что при $k > 1$ на интервале $0 < s < \frac{\pi}{2}$ появляется новое решение (почему?).

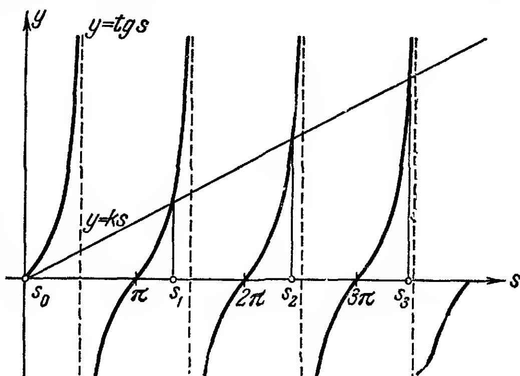


Рис. 132.

Легко получить *асимптотическое выражение* для решения s_n уравнения (3), пригодное для больших n. Пусть для определенности $k < 1$. Тогда из рис. 132 получаем искомое выражение $s_n = n\pi + \frac{\pi}{2} - \alpha_n$, где $\alpha_n \to 0$ при $n \to \infty$; в других обозначениях (п. III.11): $s_n = n\pi + \frac{\pi}{2} + o(1)$. Если мы хотим уточнить это разложение, то надо его подставить в (3), что даст

$$\operatorname{tg}\left(n\pi + \frac{\pi}{2} - \alpha_n\right) = k\left(n\pi + \frac{\pi}{2} - \alpha_n\right)$$

и после преобразований

$$\cos\alpha_n = k\left(n\pi + \frac{\pi}{2} - \alpha_n\right)\sin\alpha_n. \tag{4}$$

Отсюда

$$\alpha_n \sim \sin\alpha_n = \frac{\cos\alpha_n}{k\left(n\pi + \frac{\pi}{2} - \alpha_n\right)} \sim \frac{1}{k\pi n}, \text{ т. е. } \alpha_n = \frac{1}{k\pi n} + o\left(\frac{1}{n}\right).$$

Если нужно дальнейшее уточнение, то можно, например, в формуле (4) обозначить $\frac{1}{n} = t \to 0$, $\alpha_n = \alpha(t) \underset{t \to 0}{\longrightarrow} 0$, что даст

$$t\cos\alpha = k\left[\pi + \left(\frac{\pi}{2} - \alpha\right)t\right]\sin\alpha \qquad (\alpha = \alpha(t);\ \alpha(0) = 0),$$

после чего написать первые члены разложения $\alpha(t)$ в ряд Маклорена (вида (IV.54), но по степеням t). Вычисления, которые мы предоставляем читателю, дают

$$\alpha = \frac{1}{k\pi}t - \frac{1}{2k\pi}t^2 + \frac{1}{k\pi}\left(\frac{1}{4} + \frac{1}{k\pi^2} - \frac{1}{3k^2\pi^2}\right)t^3 + \ldots = \frac{1}{k\pi n} - \frac{1}{2k\pi n^2} + \ldots$$

Применение формулы (IV.60) дает асимптотическое выражение для положительных решений уравнения (2) при больших n:

$$x_n=\sqrt{\frac{s_n}{a}}=a^{-1/2}\sqrt{n\pi+\frac{\pi}{2}-\frac{1}{k\pi n}+\frac{1}{2k\pi n^2}-\ldots}=$$
$$=\left(\frac{n\pi}{a}\right)^{1/2}\left(1+\frac{1}{2n}-\frac{1}{k\pi^2 n^2}+\ldots\right)^{1/2}=$$
$$=\left(\frac{n\pi}{a}\right)^{1/2}\left[1+\frac{1}{4n}-\left(\frac{1}{2k\pi^2}+\frac{1}{32}\right)\frac{1}{n^2}+\ldots\right].$$

2. Методы проб, хорд и касательных. *Метод проб*, с которого часто начинают, состоит в следующем. Пусть для определенности $f(a)<0; f(b)>0$. Тогда берут произвольно значение c между a и b и вычисляют $f(c)$, причем тут существен *только знак* $f(c)$. Допустим, что получится $f(c)>0$. Это значит, что произошел «перелет», следовательно, $a<\alpha<c$. Тогда берут какое-либо значение d между a и c, вычисляют $f(d)$; если $f(d)<0$, то произошел «недолет», т. е. $d<\alpha<c$, и т. д. При этом значения c, d, ... берутся более или менее произвольными, удобными для вычисления; правда, если, например, $|f(a)|$ значительно меньше, чем $f(b)$, то довольно вероятно, что α окажется ближе к a, чем к b, и поэтому c следует взять поближе к a и т. п.

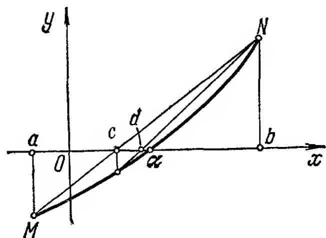


Рис. 133.

Метод хорд состоит в том, что в качестве c берется не произвольная точка, а (рис. 133) точка пересечения оси x с хордой графика, проведенной через точки $M[a; f(a)]$ и $N[b; f(b)]$. Другими словами, мы как бы приближенно принимаем дугу графика за отрезок прямой, т. е. производим *линейную интерполяцию*, что является достаточно обоснованным, если интервал a, b не слишком велик. Для нахождения точки c напишем уравнение хорды MN [см. уравнение (II.22)]:

$$\frac{y-f(a)}{f(b)-f(a)}=\frac{x-a}{b-a},$$

а затем, положив $y=0$, найдем соответствующее значение $x=c$:

$$c=a-\frac{f(a)(b-a)}{f(b)-f(a)}=b-\frac{f(b)(b-a)}{f(b)-f(a)}. \tag{5}$$

Если необходимо, это построение можно повторить (см. рис. 133).

В *методе касательных* (он же называется *методом Ньютона*) за c берется точка пересечения оси x с касательной, проведенной к графику в одном из концов рассматриваемой дуги. Уравнение касатель-

ной, изображенной на рис. 134, имеет вид [см. уравнение (IV.5)]

$$y-f(b)=f'(b)(x-b),$$

откуда, полагая $y=0$, найдем

$$c=b-\frac{f(b)}{f'(b)}. \tag{6}$$

И здесь построение можно повторить (см. рис. 134).

Метод Ньютона можно истолковать независимо от его геометрического смысла. Обозначим нулевое приближение решения через x_0 и разложим левую часть (1) по степеням $x-x_0$ в силу формулы Тейлора (IV.53); мы получим уравнение

$$f(x_0)+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\ldots=0.$$

Если произвести линеаризацию, т. е. отбросить члены выше первого порядка малости, получим *линеаризованное уравнение* (1):

$$f(x_0)+f'(x_0)(x-x_0)=0.$$

Решение этого линеаризованного уравнения

$$x_1=x_0-\frac{f(x_0)}{f'(x_0)}$$

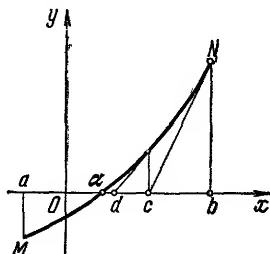


Рис. 134.

можно принять за первое приближение решения уравнения (1); мы приходим к той же формуле (6). Из первого приближения можно получить второе по формуле

$$x_2=x_1-\frac{f(x_1)}{f'(x_1)} \tag{7}$$

и т. д. Метод Ньютона всегда приводит к цели, если только нулевое приближение не лежит слишком далеко от искомого решения.

Иногда применяется следующий вариант метода Ньютона: знаменатель формулы (7), а также формул для дальнейших приближений заменяют на $f'(x_0)$; геометрически это означает, что все наклонные прямые на рис. 134 проводят параллельно касательной в исходной точке N. Метод в этом варианте сходится несколько хуже, но подсчет каждого приближения, естественно, упрощается.

Комбинированный метод основан на том соображении, что если рассматриваемый участок графика не имеет ни изломов, ни точек перегиба, то метод хорд и метод касательных дают точки, расположенные по разные стороны от искомого корня. Если, например, график расположен, как на рис. 135, то, отправляясь от интервала a, b, можно построить точку a_1 по методу хорд, а точку b_1 по методу касательных, в результате чего получится новый интервал a_1, b_1, на котором лежит искомый корень α. Проделав аналогичное построение на интервале a_1, b_1, получим новый интервал a_2, b_2, содержащий искомый корень, и т. д. При этом получается *двустороннее приближение* к этому корню, которое обрывается при достижении требуемой точности.

Рассмотрим, например, уравнение

$$x^3+x^2-3=0, \tag{8}$$

коэффициенты которого будем считать совершенно точными. Исследование производных показывает (проверьте!), что при $-\infty<x<-\frac{2}{3}$ левая часть,

которую мы обозначим через $f(x)$, возрастает от $-\infty$ до $-2\frac{23}{27}$, затем при $-\frac{2}{3}<x<0$ убывает до -3 и далее возрастает до ∞ и имеет единственную точку перегиба при $x=-\frac{1}{3}$. Значит, уравнение имеет единственный вещественный и притом положительный корень α. Так как $f(0)=-3$, $f(1)=-1$, $f(2)=9$ (рис. 136), то $1<\alpha<2$.

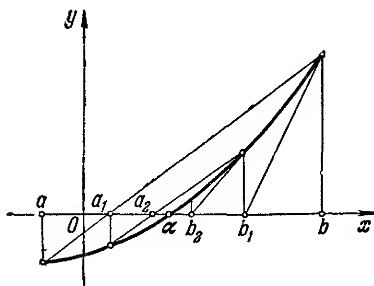


Рис. 135.

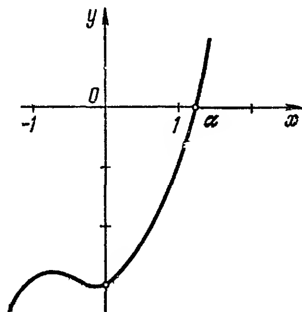


Рис. 136.

Согласно методу проб вычисляем $f(1,1)=-0,459$; $f(1,2)=0,168$, т. е. $1,1<\alpha<1,2$ (грубая *«прикидка»* корня проводится с помощью метода проб). Полагая $a=1,1$; $b=1,2$, применяем формулы (5) и (6) согласно комбинированному методу:

$$a_1=1,2-\frac{0,168\cdot 0,1}{0,168+0,459}=1,173,$$

$$b_1=1,2-\frac{0,168}{6,72}=1,175.$$

Таким образом, с точностью до 0,001 можно положить $\alpha=1,174$. Если эта точность недостаточна, то можно провести дальнейшее вычисление: $f(1,174)=-0,003628$ («недолет»; вычисление с точностью до 10^{-6}); $f(1,175)=0,002859$. Приняв $a=1,174$, $b=1,175$, получаем по комбинированному методу после вычислений с точностью до 10^{-7}: $a_2=1,1745593$; $b_2=1,1745596$. Таким образом, с точностью до 0,000001 можно положить $\alpha=1,174559$. Как быстро увеличивается степень точности!

3. Метод итераций. Методы, описанные в п. 2, принадлежат к числу итерационных методов (иначе говоря, методов последовательных приближений), в которых некоторый единообразный процесс последовательно повторяется (*«итерируется»*, от латинского «итерацио» — повторение), в результате чего получаются все более точные приближенные решения. Это единообразие имеет многочисленные удобства, в частности, в применении быстродействующих вычислительных машин.

В общем виде в применении к уравнению (1) метод итераций выглядит так: уравнение переписывается в равносильной форме:

$$x=\varphi(x). \tag{9}$$

Затем выбирается некоторое значение $x=x_0$ в качестве нулевого приближения; желательно, чтобы оно было по возможности ближе к искомому решению, если о последнем что-либо известно. Последующие приближения вычисляются по формулам $x_1=\varphi(x_0)$, $x_2=\varphi(x_1), \ldots,$ вообще

$$x_{n+1}=\varphi(x_n). \tag{10}$$

При этом может быть два случая.

1) *Процесс может сходиться*, т. е. последовательные приближения x_n при $n\to\infty$ стремятся к некоторому пределу $\bar{x}$; в этом случае, переходя в формуле (10) к пределу при $n\to\infty$, видим, что $x=\bar{x}$ является решением уравнения (9).

2) *Процесс может расходиться*, т. е. конечного предела построенных «приближений» существовать не будет. Из этого отнюдь не следует, что и решения уравнения (9) не существует, просто могло оказаться, что процесс итераций выбран неудачно. (Впрочем, и в случае сходимости бывает, что в пределе получается не то решение, около которого мы выбрали x_0, а другое, быть может, даже не имеющее физического смысла.)

Поясним сказанное на простом примере уравнения, которое можно решить без всякой «науки»,

$$x=\frac{x}{2}+1 \tag{11}$$

с очевидным решением $\bar{x}=2$. Если положить $x_0=0$ и вычислять с точностью до 0,001, то получим $x_1=1{,}000$; $x_2=1{,}500$; $x_3=1{,}750$; $x_4=1{,}875$; $x_5=1{,}938$; $x_6=1{,}969$; $x_7=1{,}984$; $x_8=1{,}992$; $x_9=1{,}996$; $x_{10}=1{,}998$; $x_{11}=1{,}999$; $x_{12}=2{,}000$; $x_{13}=2{,}000$, т. е. процесс практически сошелся.

Если взамен (11) рассмотреть уравнение

$$x=\frac{x}{10}+1$$

и принять $x_0=0$, то с точностью до 0,001 будет $x_1=1{,}000$; $x_2=1{,}100$; $x_3=1{,}110$; $x_4=1{,}111$; $x_5=1{,}111$, т. е. процесс практически сошелся уже после четырех итераций.

Если уравнение (11) разрешить относительно x, стоящего в правой части, т. е. переписать в равносильной форме

$$x=2x-2, \tag{12}$$

и начать с $x_0=0$, то мы получим последовательность $x_1=-2$, $x_2=-6$, $x_3=-14$ и т. д., т. е. процесс сходиться не будет. Это можно было предвидеть, заметив, что из (10) вытекает равенство

$$x_{n+1}-x_n=\varphi(x_n)-\varphi(x_{n-1}), \tag{13}$$

т. е. $x_2-x_1=\varphi(x_1)-\varphi(x_0)$; $x_3-x_2=\varphi(x_2)-\varphi(x_1)$ и т. д. Если значения функции меняются медленнее, чем значения аргумента,

точнее, если

$$|\varphi(x)-\varphi(\tilde{x})| \leqslant k|x-\tilde{x}| \qquad (k=\mathrm{const}<1), \tag{14}$$

то расстояния между последовательными приближениями будут быстро стремиться к нулю и процесс итераций сходится, притом тем быстрее, чем меньше k. Неравенство (14) должно выполняться для всех x, $\tilde{x}$ либо, во всяком случае, вблизи искомого корня $\bar{x}$ уравнения (9). В п. 4 будет показано, что неравенство (14) выполняется, если $|\varphi'(x)| \leqslant k$.

Мы видим, что уравнения (11) и (12) полностью равносильны, но порождают различные итерационные процессы. И в других случаях уравнение (1) можно переписать в форме (9) многими способами, каждый из которых порождает свой итерационный процесс, причем одни из этих процессов могут оказаться быстро сходящимися и потому наиболее удобными, другие — медленно сходящимися, а третьи — даже вовсе расходящимися. В частности, легко проверить, что если уравнение (1) записать в виде

$$x=x-\frac{f(x)(b-x)}{f(b)-f(x)},$$

то, если начинать с $x_0=a$ [см. формулу (5)], получится метод хорд, а если уравнение (1) записать в виде

$$x=x-\frac{f(x)}{f'(x)}, \tag{15}$$

то получится метод касательных.

В более сложных примерах, чем были разобраны выше, чаще всего не проводят теоретического доказательства сходимости процесса итераций, а просто вычисляют несколько приближений и по их виду делают вывод о сходимости или расходимости процесса. Если сочтут, что какое-либо приближение достаточно мало отличается от предыдущего, например, если отличие выходит за рамки принятой степени точности, то процесс итераций обрывают. Во всяком случае, это свидетельствует о том, что достигнутое приближение удовлетворяет уравнению (9) с хорошей точностью, так как если $|x_n-x_{n+1}|<h$, то и $|x_n-\varphi(x_n)|<h$.

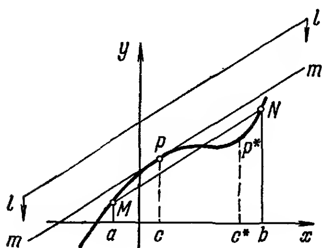


Рис. 137.

4. Формула конечных приращений. Неравенство (14) можно проверить с помощью так называемой формулы конечных приращений, которую мы сейчас выведем. Допустим, что функция $y=\varphi(x)$ на интервале a, b (включая концы) непрерывна вместе со своей производной. Рассмотрим (рис. 137) график этой функции на интервале $a \leqslant x \leqslant b$, проведем хорду MN, стягивающую его концы, и допустим для определенности, что график хотя бы частично расположен над этой хордой.

Проведем тогда выше графика прямую $ll \parallel MN$ и станем ее непрерывно опускать, оставляя параллельной MN. Тогда в некотором

положении она коснется графика в точке P, т. е. *на гладкой дуге непременно найдется по крайней мере одна точка, в которой касательная параллельна хорде, стягивающей эту дугу.* Если приравнять угловые коэффициенты хорды и касательной, то мы получим

$$\frac{\varphi(b)-\varphi(a)}{b-a}=\varphi'(c), \quad \text{т. е.} \quad \varphi(b)-\varphi(a)=\varphi'(c)(b-a), \qquad (16)$$

где c — некоторая точка между a и b. Формула (16) называется *формулой конечных приращений* (так как расстояние a от b может не быть малым) или *теоремой Лагранжа* по имени выдающегося французского математика и механика Ж. Лагра́нжа (1736—1813). Отметим, что значение c, фигурирующее в формуле (16), для данной функции и данного интервала a, b никак не является произвольным, хотя для c может получиться несколько пригодных значений. Например, для рис. 137 в формуле (16) вместо c можно взять c^*, так как в точке P^* касательная также параллельна хорде MN. При применении формулы (16) значение c обычно неизвестно, однако о c часто достаточно знать, что оно находится где-то между a и b.

Например, пусть дано, что на некотором интервале $|\varphi'(x)|\leqslant k$. Тогда, применяя формулу (16) к любым двум точкам x и $\tilde{x}$ этого интервала, увидим, что для них $|f(x)-f(\tilde{x})|\leqslant k|x-\tilde{x}|$ [см. формулу (14)].

Из формул (13) и (16) вытекает также, что если последовательные приближения находятся недалеко от точного решения $\bar{x}$, так что $\varphi'(x)$ меняется мало, то *процесс итераций сходится* примерно *со скоростью геометрической прогрессии* со знаменателем $\varphi'(\bar{x})$. Если бы разности между последовательными приближениями образовывали точно геометрическую прогрессию, как в примере (11), то ее первый член $a=x_1-x_0$, а знаменатель $q=\frac{x_2-x_1}{x_1-x_0}$. Поэтому сумма всей прогрессии, т. е. $\bar{x}-x_0$, равнялась бы

$$\frac{a}{1-q}=\frac{x_1-x_0}{1-\frac{x_2-x_1}{x_1-x_0}}=\frac{(x_1-x_0)^2}{2x_1-x_0-x_2},$$

откуда

$$\bar{x}=x_0+\frac{(x_1-x_0)^2}{2x_1-x_0-x_2}=\frac{x_1^2-x_0x_2}{2x_1-x_0-x_2}. \qquad (17)$$

В более сложных примерах последовательные разности лишь напоминают геометрическую прогрессию. Тогда формула (17) не дает точного решения, но дает возможность «перескочить» через несколько приближений и получить приближенное значение решения, от которого можно вновь начать итерации.

Особую роль играет итерационный процесс Ньютона. В самом деле, производная от правой части (15), т. е.

$$1-\frac{f'f'-ff''}{f'^2}=\frac{ff''}{f'^2},$$

обращается в нуль при $x=\bar{x}$, так как $f(\bar{x})=0$. Значит, в силу предыдущего *метод Ньютона сходится быстрее геометрической прогрессии с любым знаменателем*. Скорость этой сходимости легко установить на следующем простом типичном примере. Пусть рассматриваются приближения по способу Ньютона к нулевому корню уравнения $x+x^2=0$. Эти приближения связаны друг с другом соотношением

$$x_{n+1}=x_n-\frac{x_n+x_n^2}{1+2x_n}=\frac{x_n^2}{1+2x_n}\approx x_n^2.$$

Для оценки скорости сходимости заменим это приближенное равенство на точное; тогда последовательно получим $x_1=x_0^2$, $x_2=x_1^2=x_0^4$, $x_3=x_2^2=x_0^8$ и т. д., вообще $x_n=x_0^{2^n}$. При $|x_0|<1$ правая часть с увеличением n стремится к нулю быстрее любой экспоненты.

5. Метод малого параметра. *Метод малого параметра*, он же *метод возмущений*, как и метод итераций, представляет собой один из наиболее универсальных методов в математике и заключается в следующем. Пусть формулировка некоторой задачи, помимо основных неизвестных величин, содержит некоторый параметр α, причем эта задача при каком-то значении $\alpha=\alpha_0$ может быть более или менее легко решена (*невозмущенное решение*). Тогда решение задачи при α, близких к α_0 (*возмущенное решение*), во многих случаях можно получить разложенным по степеням $\alpha-\alpha_0$ с той или иной степенью точности, подобно формулам (IV.49), (IV.50), (IV.51) и т. д. При этом первый член разложения, не содержащий $\alpha-\alpha_0$, получается при $\alpha=\alpha_0$, т. е. дает невозмущенное решение. Дальнейшие же члены дают поправки на «возмущение» решения; эти поправки имеют первый, второй и т. д. порядки малости (по сравнению с $\alpha-\alpha_0$). Эти члены обычно находятся *по методу неопределенных коэффициентов*, т. е. коэффициенты при $(\alpha-\alpha_0)$, $(\alpha-\alpha_0)^2$ и т. д. обозначаются какими-то буквами, которые находятся затем из условий задачи. Этот метод дает хороший результат только при α, близких к α_0, при этом чем $|\alpha-\alpha_0|$ меньше, тем меньше членов разложения нужно вычислять; так как часто принимают $\alpha_0=0$, то отсюда и происходит название метода. Сколько именно брать членов, можно определять так, как описано в конце п. III.6. Следует иметь также в виду, что при больших $|\alpha-\alpha_0|$ метод может привести к принципиальным ошибкам, так как может получиться, что отбрасываемые члены более существенны, чем оставляемые.

Таким образом, метод малого параметра дает возможность, исходя из решения некоторых «узловых» задач, получить решение задач, формулировка которых близка к этим узловым, если, конечно, изменение формулировки не влечет за собой принципиального, качественного изменения решения. Во многих задачах уже вид первого члена, содержащего параметр, дает возможность сделать полезные выводы о зависимости решения от параметра при его малом изменении.

Пример. Найдем решение уравнения

$$x^3-\alpha x^2+1=0 \tag{18}$$

при малых $|\alpha|$ с точностью до величин поряда α^3 включительно. Для этого заметим, что при $\alpha=0$ получается уравнение $x^3+1=0$ с очевидным решением $x_0=-1$. Поэтому пишем

$$x_\alpha=-1+a\alpha+b\alpha^2+c\alpha^3+\text{члены высшего порядка малости.}$$

Подставляя это выражение в (18) и выписывая члены только до α^3, получим (проверьте!)

$$[-1+3a\alpha+3b\alpha^2-3a^2\alpha^2-6ab\alpha^3+3c\alpha^3+a^3\alpha^3]-$$
$$-\alpha(1-2a\alpha-2b\alpha^2+a^2\alpha^2)+1+\text{члены высшего порядка малости}=0.$$

Отсюда, приравнивая коэффициенты при одинаковых степенях α, получим $3a-1=0$, $3b-3a^2+2a=0$, $-6ab+3c+a^3+2b-a^2=0$ и последовательно находим $a=\frac{1}{3}$, $b=-\frac{1}{9}$, $c=\frac{2}{81}$, т. е. решение уравнения (18)

$$x_\alpha=-1+\frac{\alpha}{3}-\frac{\alpha^2}{9}+\frac{2\alpha^3}{81} \tag{19}$$

с точностью до величин высшего порядка малости относительно α.

Тот же результат можно получить, *применяя непосредственно формулу Тейлора* (IV.51) с измененными обозначениями:

$$x_\alpha=x_0+\left(\frac{dx}{d\alpha}\right)_0\alpha+\frac{1}{2!}\left(\frac{d^2x}{d\alpha^2}\right)_0\alpha^2+\frac{1}{3!}\left(\frac{d^3x}{d\alpha^3}\right)_0\alpha^3; \tag{20}$$

индекс «нуль» указывает на подстановку $\alpha=0$. Для этого дифференцируем равенство (18) по α, как в п. IV.11:

$$3x^2\frac{dx}{d\alpha}-x^2-2\alpha x\frac{dx}{d\alpha}=0,$$

$$6x\left(\frac{dx}{d\alpha}\right)^2+3x^2\frac{d^2x}{d\alpha^2}-4x\frac{dx}{d\alpha}-2\alpha\left(\frac{dx}{d\alpha}\right)^2-2\alpha x\frac{d^2x}{d\alpha^2}=0,$$

$$6\left(\frac{dx}{d\alpha}\right)^3+18x\frac{dx}{d\alpha}\frac{d^2x}{d\alpha^2}+3x^2\frac{d^3x}{d\alpha^3}-6\left(\frac{dx}{d\alpha}\right)^2-6x\frac{d^2x}{d\alpha^2}-6\alpha\frac{dx}{d\alpha}\frac{d^2x}{d\alpha^2}-2\alpha x\frac{d^3x}{d\alpha^3}=0.$$

Подставляя $\alpha=0$, $x=-1$, получим

$$3\left(\frac{dx}{d\alpha}\right)_0-1=0,\quad -6\left(\frac{dx}{d\alpha}\right)_0^2+3\left(\frac{d^2x}{d\alpha^2}\right)_0+4\left(\frac{dx}{d\alpha}\right)_0=0,$$

$$6\left(\frac{dx}{d\alpha}\right)_0^3-18\left(\frac{dx}{d\alpha}\right)_0\left(\frac{d^2x}{d\alpha^2}\right)_0+3\left(\frac{d^3x}{d\alpha^3}\right)_0-6\left(\frac{dx}{d\alpha}\right)_0^2+6\left(\frac{d^2x}{d\alpha^2}\right)_0=0,$$

откуда получаем последовательно

$$\left(\frac{dx}{d\alpha}\right)_0=\frac{1}{3};\qquad \left(\frac{d^2x}{d\alpha^2}\right)_0=-\frac{2}{9};\qquad \left(\frac{d^3x}{d\alpha^3}\right)_0=\frac{4}{27}.$$

Отсюда из формулы (20) вытекает разложение (19), дающее хорошую точность при малых $|\alpha|$.

Метод малого параметра непосредственно связан с методом итераций п. 3. что мы продемонстрируем на том же примере (18). Удобно, чтобы невозмущенное решение было нулевым; это достигается с помощью подстановки $x=-1+y$, откуда

$$-1+3y-3y^2+y^3-\alpha+2\alpha y-\alpha y^2+1=0,$$

т. е.

$$y=\frac{1}{3}\alpha-\frac{2}{3}\alpha y+y^2+\frac{1}{3}\alpha y^2-\frac{1}{3}y^3.$$

Если теперь проводить итерации, начиная от значения $y_0=0$ и отбрасывая в разложении члены выше третьего порядка малости, то за три шага мы придем к требуемому разложению. При этом легко проверить, что в каждом приближении можно отбрасывать члены, порядок малости которых выше номера приближения.

§ 2. Интерполяция

6. Интерполяционная формула Лагранжа. В п. I.22 мы познакомились с линейной интерполяцией, которая состоит в приближенной замене рассматриваемой функции $y=f(x)$ линейной функцией $y=ax+b$, совпадающей с $f(x)$ в некоторых двух точках. Очевидно, что если вместо линейной функции использовать многочлен n-й степени

$$P(x)=P_n(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n,$$

то точность такой замены можно повысить.

Так как многочлен $P_n(x)$, *аппроксимирующий* (приближающий, приближенно заменяющий) функцию $f(x)$, содержит $n+1$ параметров, которыми служат коэффициенты, то при его подборе можно поставить $n+1$ условий. Рассмотрим для простоты случай многочлена второй степени

$$P(x)=P_2(x)=ax^2+bx+c$$

(общий случай разбирается аналогично); при его подборе можно поставить три условия. Часто *требуют, чтобы этот многочлен совпадал с функцией f в некоторых трех заданных точках*:

$$P(x_1)=f(x_1);\quad P(x_2)=f(x_2);\quad P(x_3)=f(x_3). \tag{21}$$

Эти значения также считаются известными.

Прежде всего ясно, что искомый многочлен *может быть только один*: если бы другой многочлен второй степени $Q(x)$ удовлетворял условиям (21), то разность $P(x)-Q(x)$ — также многочлен второй степени — равнялась бы нулю при $x=x_1$; $x=x_2$; $x=x_3$. Но если многочлен второй степени равен нулю в трех точках, то он равен нулю тождественно, т. е. все коэффициенты равны нулю (почему?); итак, $Q(x)\equiv P(x)$.

Лагранж предложил искать многочлен $P(x)$ в форме

$$P(x)=A(x-x_2)(x-x_3)+B(x-x_1)(x-x_3)+ \\ +C(x-x_1)(x-x_2), \tag{22}$$

где A, B, C — постоянные, пока неизвестные. Ясно, что это — многочлен второй степени. Для выбора постоянных A, B, C воспользуемся условиями (21), заметив, что при подстановке $x=x_1$, x_2 или x_3 в правой части формулы (22) остается лишь одно слагаемое. Получим

$$f(x_1)=A(x_1-x_2)(x_1-x_3),\quad f(x_2)=B(x_2-x_1)(x_2-x_3),$$

$$f(x_3)=C(x_3-x_1)(x_3-x_2).$$

Найдя отсюда A, B, C и подставляя их в (22), получим *интерполяционную формулу Лагранжа*

$$f(x) \approx P_2(x) = f(x_1)\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + \\ + f(x_2)\frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + f(x_3)\frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}. \quad (23)$$

При применении этой формулы желательно, чтобы ни одна из разностей x_1-x_2, x_1-x_3, x_2-x_3 не была чрезмерно малой (почему?).

Взамен (21) можно поставить, например, такие три условия:

$$P(x_1)=f(x_1); \quad P'(x_1)=f'(x_1); \quad P(x_2)=f(x_2).$$

Тогда многочлен $P(x)$ вместо (22) можно искать в виде

$$P(x)=A(x-x_2)(x-2x_1+x_2)+B(x-x_1)(x-x_2)+C(x-x_1)^2.$$

(Найдите коэффициенты A, B, C!)

7. Конечные разности и их связь с производными. Прежде чем перейти к дальнейшему, рассмотрим одно из важных понятий современной математики, понятие *конечной разности*. Пусть $y=f(x)$; тогда при данном h выражение

$$\Delta_h y = f(x+h)-f(x)$$

называется *конечной разностью первого порядка* («*первой разностью*») *функции* f *с шагом* h, а выражение

$$\frac{1}{h}\Delta_h y = \frac{f(x+h)-f(x)}{h}$$

— *первой разделенной разностью*. Из определения производной (п. IV.2) ясно, что *при достаточно малом* h будет

$$\frac{1}{h}\Delta_h y \approx y', \quad (24)$$

точнее,

$$y' = \lim_{h\to 0}\frac{1}{h}\Delta_h y.$$

Пусть, например, $y=x^3$. Тогда

$$\Delta_h y = (x+h)^3 - x^3 = 3x^2h + 3xh^2 + h^3,$$

$$\frac{1}{h}\Delta_h y = 3x^2 + 3xh + h^2,$$

$$\lim_{h\to 0}\left(\frac{1}{h}\Delta_h y\right) = \lim_{h\to 0}(3x^2+3xh+h^2) = 3x^2 = y'.$$

Из очевидных свойств разностей отметим такие:

$$\Delta_h(y_1+y_2)=\Delta_h y_1+\Delta_h y_2; \quad \Delta_h(Cy)=C\Delta_h y \quad (C=\text{const}).$$

От разности можно взять разность (*вторую*):

$$\Delta_h^2 y = \Delta_h(\Delta_h y) = \Delta_h[f(x+h)-f(x)] = [f(x+2h)-f(x+h)] - [f(x+h)-f(x)] = f(x+2h)-2f(x+h)+f(x).$$

Аналогично определяется вторая разделенная разность

$$\frac{1}{h}\Delta_h\left(\frac{1}{h}\Delta_h y\right) = \frac{1}{h^2}\Delta_h(\Delta_h y) = \frac{1}{h^2}\Delta_h^2 y = \frac{f(x+2h)-2f(x+h)+f(x)}{h^2}.$$

Так как взятие разделенной разности с малым шагом примерно равносильно дифференцированию, то *вторая разделенная разность с малым шагом приближенно равна второй производной*, точнее,

$$y'' = \lim_{h\to 0}\frac{1}{h^2}\Delta_h^2 y = \lim_{h\to 0}\frac{f(x+2h)-2f(x+h)+f(x)}{h^2}. \qquad (25)$$

Так, в предыдущем примере

$$\Delta_h^2 y = \Delta_h(3x^2h+3xh^2+h^3) = 3(x+h)^2h+3(x+h)h^2+h^3-3x^2h-3xh^2-h^3 = 6xh^2+6h^3;$$

$$\lim_{h\to 0}\frac{1}{h^2}\Delta_h^2 y = \lim_{h\to 0}(6x+6h) = 6x = y''.$$

Аналогично определяются третья разность $\Delta_h^3 y = \Delta_h(\Delta_h^2 y)$ и третья разделенная разность $\frac{1}{h^3}\Delta_h^3 y$, которая в пределе дает третью производную y''', и т. д.

Вычисление этих разностей особенно удобно, если функция задана табличным способом. Для таблицы общего вида (I.2) можно написать $\Delta y_1 = y_2 - y_1$, $\Delta y_2 = y_3 - y_2$, вообще $\Delta y_k = y_{k+1} - y_k$; здесь нижний индекс у Δy_k означает уже не шаг, который для таблицы зафиксирован, а номер разности. Далее $\Delta^2 y_1 = \Delta y_2 - \Delta y_1$, $\Delta^2 y_2 = \Delta y_3 - \Delta y_2$ и т. д. Например, при $h = 0{,}1$:

x	10,0	10,1	10,2	10,3	10,4	10,5	10,6	10,7
y	1,00000	1,00432	1,00860	1,01284	1,01703	1,02119	1,02531	1,02938
$10^5\Delta y$	432	428	424	419	416	412	407	
$10^5\Delta^2 y$	−4	−4	−5	−3	−4	−5		
$10^5\Delta^3 y$	0	−1	2	−1	−1			

(Приведен отрывок из таблицы логарифмов, причем значения разностей умножены на 100 000, чтобы избавиться от десятичных нулей.)

Малость и почти постоянство вторых разностей в приведенном примере указывают на плавность хода изменения функции, на отсутствие случайных «выпадов» из этого хода. Такая закономерность может проявляться в разностях более высокого порядка и всегда свидетельствует о «правильности» хода изменения функции. Конечно, если шаг не мал, а также вблизи точек разрыва и т. п. разности могут и не быть малыми, но обычно в них проявляется та или иная закономерность. В то же время случайные ошибки в составлении таблицы резко проявляются в высших разностях и, таким образом, могут быть уловлены. Пусть, например, в приведенной таблице вместо 1,01284 по ошибке было бы написано 1,01294. Тогда четвертая строка имела бы вид (проверьте!) —4, +6, —25, +7, —4, —5, т. е. закономерность, очевидно, нарушена. По этой же причине, если таблица получена из эксперимента, проведенного не с особенно высокой точностью, то разности выше второго порядка применяются редко; часто ограничиваются даже только первыми разностями.

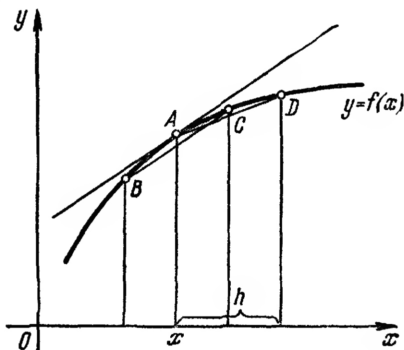


Рис. 138.

Иногда разность $y_{k+1}-y_k$ сопоставляют не значению x_k, как выше, а значению $x_k+\frac{h}{2}$, которое обозначают $x_{k+\frac{1}{2}}$ и условно называют *полуцелым*, в отличие от «*целых*» значений x_k (конечно, целыми или полуцелыми на самом деле являются номера, а не сами x). Тогда разность называется *центральной* и обозначается $\delta_{k+\frac{1}{2}}y=y_{k+1}-y_k$. При делении ее на шаг h получается разделенная центральная разность. Аналогично составляются центральные разности второго порядка $\delta_k^2 y=\delta_{k+\frac{1}{2}}y-\delta_{k-\frac{1}{2}}y$; они опять определены при «целых» значениях аргумента и т. д.

Из рис. 138 видно, что разделенная центральная разность, равная угловому коэффициенту хорды BC, ближе к производной (к угловому коэффициенту касательной в точке A), чем простая разделенная разность (угловой коэффициент хорды AD). Это легко проверить с помощью ряда Тейлора (IV.52): разность

$$\frac{\Delta y}{h}-y'=\frac{y(x+h)-y(x)}{h}-y'(x)=\frac{h}{2}y''(x)+\frac{h^2}{6}y'''(x)+\ldots$$

при малых h имеет порядок h, тогда как разность

$$\frac{\delta y}{h}-y'=\frac{y\left(x+\frac{h}{2}\right)-y\left(x-\frac{h}{2}\right)}{h}-y'(x)=\frac{h^2}{24}y'''(x)+\frac{h^4}{1920}y^{\mathrm{V}}(x)+\ldots$$

(проверьте!) имеет порядок h^2. (Аналогично находится порядок ошибки других приближенных формул при малом шаге.) Итак, приближенные значения производной лучше определять с помощью центральной разделенной разности, чем по формуле (24). В п. 9 будет указан еще более точный способ приближенного вычисления производных любого порядка.

Если разделенные разности при малом шаге близки к соответствующим производным и во многом им аналогичны, то сами (неразделенные) разности близки к соответствующим дифференциалам. Например, из формул (25) видим, что $\frac{1}{h^2}\Delta_h^2 y = y'' + \alpha$ ($|\alpha| \ll 1$, т. е. α — бесконечно малая при $h \to 0$); отсюда

$$\Delta_h^2 y = y''h^2 + \alpha h^2 = y''(\Delta x)^2 + \alpha h^2 = d^2y + \alpha h^2 \qquad (|\alpha h^2| \ll h^2).$$

Значит, при $y'' \neq 0$, $\Delta_h^2 y$ и $d^2 y$ различаются на величину высшего порядка малости и при $h \to 0$ являются бесконечно малыми эквивалентными (см. п. III.8).

Теория конечных разностей развивалась параллельно с развитием основных разделов математического анализа; впервые систематическое изложение исчисления конечных разностей было дано в 1715 г. Тейлором. В настоящее время конечные разности широко применяются в теоретических и прикладных исследованиях, особенно в связи с электронными быстродействующими математическими машинами.

8. Интерполяционные формулы Ньютона. Если расстояние h между соседними значениями x, при которых задается функция f, является постоянным, то можно пользоваться более удобными формулами, чем формула (23). Пусть, например, известны значения

$$f(x_0) = y_0; \quad f(x_1) = y_1; \quad f(x_2) = y_2; \quad f(x_3) = y_3,$$

где $x_1 = x_0 + h$, $x_2 = x_0 + 2h$, $x_3 = x_0 + 3h$. Тогда многочлен $P(x)$, принимающий те же значения при указанных значениях x, будет иметь третью степень (см. п. 6). Ньютон предложил искать его в виде

$$P(x) = A + Bs + Cs(s-h) + Ds(s-h)(s-2h), \qquad (26)$$

где $s = x - x_0$. Согласно условию должно быть

$$y_0 = P(x_0) = P|_{s=0} = A; \quad y_1 = P(x_1) = P|_{s=h} = A + Bh;$$
$$y_2 = P(x_2) = P|_{s=2h} = A + B\cdot 2h + C\cdot 2h^2;$$
$$y_3 = P(x_3) = P|_{s=3h} = A + B\cdot 3h + C\cdot 3\cdot 2h^2 + D\cdot 3\cdot 2h^3.$$

Составляя разности (п. 7) для левых и правых частей, получим

$$\Delta y_0 = Bh; \quad \Delta y_1 = Bh + C\cdot 2h^2; \quad \Delta y_2 = Bh + C\cdot 2\cdot 2h^2 + D\cdot 3\cdot 2h^3.$$

Вторично составляя разности, а затем и в третий раз, найдем

$$\Delta^2 y_0 = C\cdot 2h^2; \quad \Delta^2 y_1 = C\cdot 2h^2 + D\cdot 3\cdot 2h^2; \quad \Delta^3 y_0 = D\cdot 3\cdot 2h^2.$$

Отсюда находим $A=y_0$; $B=\frac{\Delta y_0}{h}$; $C=\frac{\Delta^2 y_0}{2!h^2}$; $D=\frac{\Delta^3 y_0}{3!h^3}$. Подставляя эти значения в (26) и заметив, что вместо x_0 можно было начинать от любого табличного значения x_k, получим *формулу Ньютона*

$$f(x)\approx P(x)=$$
$$=y_k+\Delta y_k\frac{s}{h}+\frac{\Delta^2 y_k}{2!}\frac{s}{h}\left(\frac{s}{h}-1\right)+\frac{\Delta^3 y_k}{3!}\frac{s}{h}\left(\frac{s}{h}-1\right)\left(\frac{s}{h}-2\right), \quad (27)$$

где $s=x-x_k$.

Аналогичный вид имеет формула для интерполяционных многочленов других степеней. Увеличивая эту степень, можно подобно п. IV.16 перейти к бесконечному ряду

$$f(x)=y_k+\Delta y_k\frac{s}{h}+\frac{\Delta^2 y_k}{2!}\frac{s}{h}\left(\frac{s}{h}-1\right)+$$
$$+\frac{\Delta^3 y_k}{3!}\frac{s}{h}\left(\frac{s}{h}-1\right)\left(\frac{s}{h}-2\right)+\ldots, \quad (28)$$

причем не выписанные члены содержат соответственно разности четвертого, пятого и т. д. порядков и потому имеют четвертый, пятый и т. д. порядок малости по сравнению с шагом h. На практике эту формулу, конечно, обрывают, доводя ее до места, начиная с которого слагаемыми можно пренебречь. Если шаг велик или если мы находимся вблизи от конца интервала, на котором задана функция $f(x)$, то может оказаться, что такого места достичь нельзя: тогда и формулой (28) пользоваться нельзя.

Формулы Ньютона (27) или (28) легко применять, если функция f задана таблично, так как в этом случае легко подсчитывать разности. Особенно часто они применяются *в начале таблицы* (например, если $k=0$, x_0 — первое табличное значение аргумента, а $x_0<x<x_1$). Степень интерполяционного многочлена $P(x)$ выбирают, руководствуясь значениями разностей; например, если третьи разности очень малы (как в примере п. 7), то последний член в формуле (27) мал и его можно отбросить, т. е. ограничиться многочленом второй степени. В формуле (28) можно положить и $k=0$, $x<x_0$, если $|x-x_0|$ невелико, что приведет *к экстраполяции таблицы назад*.

Подобно (28) выводится *другая формула Ньютона*:

$$f(x)=y_{k+1}-\Delta y_k\frac{t}{h}+\frac{\Delta^2 y_{k-1}}{2!}\frac{t}{h}\left(\frac{t}{h}-1\right)-$$
$$-\frac{\Delta^3 y_{k-2}}{3!}\frac{t}{h}\left(\frac{t}{h}-1\right)\left(\frac{t}{h}-2\right)+\ldots, \quad (29)$$

где $t=x_{k+1}-x$, которая применяется, в частности, *в конце таблицы*, например, если x_{k+1} — последнее табличное значение аргумента, а $x_k<x<x_{k+1}$. Эта же формула применяется для *экстраполяции таблицы вперед*.

При интерполяции в середине таблицы желательно иметь формулу, использующую в равной мере табличные значения функции как впереди, так и позади рассматриваемого значения x. Одной из таких формул служит *формула Бесселя*, которая получается, если взять полусумму правых частей (28) и (29):

$$f(x)=\frac{y_k+y_{k+1}}{2}+\Delta y_k\left(\frac{s}{h}-\frac{1}{2}\right)+ \\ +\frac{\Delta^2 y_{k-1}+\Delta^2 y_k}{2\cdot 2!}\frac{s}{h}\left(\frac{s}{h}-1\right)+\frac{\Delta^3 y_{k-1}}{3!}\frac{s}{h}\left(\frac{s}{h}-1\right)\left(\frac{s}{h}-\frac{1}{2}\right)+\ldots, \qquad (30)$$

где $s=x-x_k$. Эта формула, обладающая высокой точностью, названа по имени немецкого астронома Ф. Бесселя (1784—1846), хотя фактически она принадлежит Ньютону.

Интерполяционные формулы применяются также к задаче *обратного интерполирования*, которая состоит в отыскании значения аргумента по заданному значению функции. Будем исходить, например, из формулы (27). Приняв равенство за точное, эту формулу можно разрешить относительно второго слагаемого в правой части, что после деления на Δy_k даст

$$\frac{s}{h}=\frac{y-y_k}{\Delta y_k}-\frac{\Delta^2 y_k}{2!\,\Delta y_k}\frac{s}{h}\left(\frac{s}{h}-1\right)-\frac{\Delta^3 y_k}{3!\,\Delta y_k}\frac{s}{h}\left(\frac{s}{h}-1\right)\left(\frac{s}{h}-2\right). \qquad (31)$$

Если задано y, то для нахождения s можно применить метод последовательных приближений (п. 3). Для этого в качестве нулевого приближения можно положить $\left(\frac{s}{h}\right)_0=\frac{y-y_k}{\Delta y_k}$; подставив это значение в правую часть (31), находим $\left(\frac{s}{h}\right)_1$ и т. д. При малом h процесс хорошо сходится.

При интерполяции разрывной функции или функции с разрывной производной надо иметь в виду, что вблизи точек разрыва качество аппроксимации может значительно понизиться, так как интерполяционный многочлен разрывов не имеет. Для имитации разрыва можно значительно сближать узлы интерполяции вблизи точек разрыва, но обычно предпочитают проводить интерполяцию только на интервалах между точками разрыва.

9. Численное дифференцирование. *Численное дифференцирование* обычно производится, если функция, от которой надо найти производную, задана таблично. Это можно сделать так: согласно методам пп. 6 или 8 заменить рассматриваемую функцию многочленом, от которого и найти производную. Например, из формулы (27) получается (проверьте!)

$$f'(x)\approx\frac{\Delta y_k}{h}+\frac{\Delta^2 y_k}{h}\left(\frac{s}{h}-\frac{1}{2}\right)+\frac{\Delta^3 y_k}{2h}\left[\left(\frac{s}{h}\right)^2-2\left(\frac{s}{h}\right)+\frac{2}{3}\right];$$

взяв в формуле (28) больше членов, можно получить более точный результат. В частности, полагая $x=x_k$ (т. е. $s=0$), получим

$$f'(x_k)\approx\frac{1}{h}\left(\Delta y_k-\frac{\Delta^2 y_k}{2}+\frac{\Delta^3 y_k}{3}\right).$$

Более точная формула имеет вид бесконечного ряда

$$f'(x_k)=\frac{1}{h}\left(\frac{\Delta y_k}{1}-\frac{\Delta^2 y_k}{2}+\frac{\Delta^3 y_k}{3}-\frac{\Delta^4 y_k}{4}+\ldots\right). \qquad (32)$$

Подобным образом можно получить формулы для производных второго и последующих порядков. Можно воспользоваться и формулами (29) и (30), а также другими интерполяционными формулами.

Отметим, в частности, формулу, вытекающую из формулы (30),

$$f'(x_k)=\frac{1}{2h}\left\{(\Delta y_{k-1}+\Delta y_k)-\frac{1}{6}(\Delta^3 y_{k-2}+\Delta^3 y_{k-1})+\right.$$
$$\left.+\frac{1}{30}(\Delta^5 y_{k-3}+\Delta^5 y_{k-2})-\ldots\right\}.$$

В ней последовательные члены имеют первый, третий, пятый и т. д. порядки малости и коэффициенты убывают быстрее, чем в ряде (32).

Если таблица функции получилась в результате эксперимента, то малая ошибка в значении функции после деления на малый шаг может привести к конечной и даже большой ошибке в значении производной. Еще хуже обстоит дело при вычислении производных высших порядков. Поэтому желательно, чтобы шаг таблицы был по крайней мере на порядок (т. е., скажем, в 10 раз) больше, чем возможная ошибка в значении функции; для вычисления производной второго порядка шаг должен быть на два порядка больше этой ошибки. Из-за этой трудности при дифференцировании эмпирических функций часто предпочитают пользоваться не интерполяционными, а другими эмпирическими формулами (ср. п. I.30), которые составляются с учетом всех опытных данных и являются значительно более устойчивыми относительно случайных ошибок эксперимента.

Интерполяционные формулы и формулы численного дифференцирования подробно разбираются в курсах приближенных вычислений.

ОПРЕДЕЛИТЕЛИ И СИСТЕМЫ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

§ 1. Определители

1. Определение. К понятию определителя мы приходим, рассматривая системы алгебраических уравнений первой степени. Рассмотрим сначала систему уравнений

$$\left.\begin{aligned} a_1x+b_1y=d_1, \\ a_2x+b_2y=d_2 \end{aligned}\right\} \qquad (1)$$

с неизвестными x и y. Решая ее, что мы предоставим сделать читателю, получим ответ

$$x=\frac{d_1b_2-b_1d_2}{a_1b_2-b_1a_2}; \quad y=\frac{a_1d_2-d_1a_2}{a_1b_2-b_1a_2}. \qquad (2)$$

Выражение $a_1b_2-b_1a_2$ называется *определителем* (*детерминантом*) *второго порядка* и записывается в виде

$$a_1b_2-b_1a_2=\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \qquad (3)$$

где вертикальные черточки — знак определителя. С помощью этого обозначения формулы (2) можно переписать в виде

$$x=\frac{\begin{vmatrix} d_1 & b_1 \\ d_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}; \quad y=\frac{\begin{vmatrix} a_1 & d_1 \\ a_2 & d_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \qquad (4)$$

Покажем на примере вычисление определителя:

$$\begin{vmatrix} 0 & -3 \\ 2 & 1 \end{vmatrix}=0\cdot 1-(-3)\,2=0+6=6.$$

Аналогичное решение системы уравнений

$$\left.\begin{aligned} a_1x+b_1y+c_1z=d_1, \\ a_2x+b_2y+c_2z=d_2, \\ a_3x+b_3y+c_3z=d_3 \end{aligned}\right\} \qquad (5)$$

приводит, взамен (2), к дробям, у которых в знаменателе стоит

$$a_1b_2c_3 - a_1c_2b_3 - b_1a_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3. \qquad (6)$$

Это выражение называется *определителем третьего порядка* и обозначается

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}. \qquad (7)$$

Преобразование выражения (6) и применение обозначения (3) приводит к формуле

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1(b_2c_3 - c_2b_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3) =$$

$$= a_1\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}, \qquad (8)$$

по которой можно вычислять значение определителя. Например,

$$\begin{vmatrix} 1 & 0 & -2 \\ -1 & 1 & \frac{1}{2} \\ 3 & 1 & 2 \end{vmatrix} = 1\begin{vmatrix} 1 & \frac{1}{2} \\ 1 & 2 \end{vmatrix} - 0\begin{vmatrix} -1 & \frac{1}{2} \\ 3 & 2 \end{vmatrix} + (-2)\begin{vmatrix} -1 & 1 \\ 3 & 1 \end{vmatrix} =$$

$$= 1\left(1\cdot 2 - \frac{1}{2}\cdot 1\right) - 2(-1\cdot 1 - 1\cdot 3) = \frac{3}{2} + 8 = 9\frac{1}{2}.$$

По аналогии с формулой (8) определяются *детерминанты* (определители) *четвертого порядка:*

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix} = a_1\begin{vmatrix} b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \\ b_4 & c_4 & d_4 \end{vmatrix} - b_1\begin{vmatrix} a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \\ a_4 & c_4 & d_4 \end{vmatrix} +$$

$$+ c_1\begin{vmatrix} a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \\ a_4 & b_4 & d_4 \end{vmatrix} - d_1\begin{vmatrix} a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \end{vmatrix}$$

(надо как следует продумать структуру выражения, стоящего в правой части), *пятого порядка* и т. д.

2. Свойства. Мы будем излагать свойства определителей на примере определителей третьего порядка (7), хотя все эти свойства справедливы для определителей любого порядка.

Определитель третьего порядка (7) имеет три *строчки* (горизонтальных ряда), три *столбца* (вертикальных ряда), девять *элементов* (чисел $a_1, b_1, \ldots, c_3$).

1. *Если переставить два параллельных ряда*, т. е. две строчки или два столбца, *то определитель умножится на* -1. Например,

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = -\begin{vmatrix} c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \\ c_3 & b_3 & a_3 \end{vmatrix} \tag{9}$$

(переставили третий и первый столбцы). Это доказывается проверкой, т. е. сравнением обеих частей по формуле (8):

$$-\begin{vmatrix} c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \\ c_3 & b_3 & a_3 \end{vmatrix} = -c_1\begin{vmatrix} b_2 & a_2 \\ b_3 & a_3 \end{vmatrix} + b_1\begin{vmatrix} c_2 & a_2 \\ c_3 & a_3 \end{vmatrix} - a_1\begin{vmatrix} c_2 & b_2 \\ c_3 & b_3 \end{vmatrix} =$$

$$= -c_1(b_2a_3 - a_2b_3) + b_1(c_2a_3 - a_2c_3) - a_1(c_2b_3 - b_2c_3).$$

Если раскрыть скобки, то легко проверить, что получится как раз выражение (6), но с переставленными слагаемыми. Этим и доказывается формула (9).

2. *Если определитель имеет два одинаковых ряда, то он равен нулю.* Например,

$$P = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0$$

(здесь одинаковые вторая и третья строчки). Действительно, если переставить эти два ряда, то по свойству 1 должно получиться $-P$, т. е. $P = -P$, откуда $P = 0$.

3. *Общий множитель, содержащийся во всех элементах одного ряда, можно вынести за знак определителя.* Например,

$$\begin{vmatrix} a_1 & kb_1 & c_1 \\ a_2 & kb_2 & c_2 \\ a_3 & kb_3 & c_3 \end{vmatrix} = k\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}. \tag{10}$$

Доказывается проверкой, как и свойство 1.

4. *Определитель, имеющий нулевой ряд, равен нулю.* Для доказательства достаточно в формуле (10) положить $k = 0$.

5. *Если все элементы какого-либо ряда*, например, второй строки, *представлены в виде суммы двух слагаемых, то весь определитель можно представить в виде суммы двух определителей по формуле*

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2' + a_2'' & b_2' + b_2'' & c_2' + c_2'' \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2' & b_2' & c_2' \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2' & b_2'' & c_2'' \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Доказывается проверкой.

6. *Если к каждому из элементов какого-нибудь ряда прибавить числа, пропорциональные соответствующим элементам какого-нибудь другого ряда, параллельного первому, то значение определителя не изменится.* Действительно, например,

$$\begin{vmatrix} a_1+kc_1 & b_1 & c_1 \\ a_2+kc_2 & b_2 & c_2 \\ a_3+kc_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} kc_1 & b_1 & c_1 \\ kc_2 & b_2 & c_2 \\ kc_3 & b_3 & c_3 \end{vmatrix} =$$

$$= \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + k \begin{vmatrix} c_1 & b_1 & c_1 \\ c_2 & b_2 & c_2 \\ c_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

(мы последовательно применяли свойства 5, 3 и 2).

7. *Определитель не меняется, если заменить его строки столбцами и обратно*, т. е.

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

(это называется *транспонированием* определителя). Доказывается проверкой.

3. Разложение определителя по элементам ряда. Нам потребуется понятие об *алгебраическом дополнении* какого-нибудь элемента в определителе. Им называется *определитель, полученный из данного вычеркиванием строчки и столбца, в которых находится данный элемент.* Но этот определитель меньшего порядка *надо взять со знаком + или —, в зависимости от положения данного элемента в исходном определителе*: для левого верхнего элемента берется +, а для других — в шахматном порядке по схеме

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

Мы будем обозначать алгебраические дополнения большими буквами. Так, в определителе (7) алгебраическое дополнение A_1 элемента a_1 равно $\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}$, алгебраическое дополнение C_2 элемента c_2 равно $-\begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix}$ и т. д.

Имеет место следующее свойство определителей, которое доказывается проверкой, как и свойства п. 2: *определитель равен сумме произведений элементов какого-нибудь из рядов на алгебраические*

дополнения этих элементов. Например,

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = b_1B_1 + b_2B_2 + b_3B_3 =$$

$$= -b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + b_2\begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix} - b_3\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}.$$

Это преобразование называется *разложением определителя* по элементам данного ряда (в приведенном примере — по элементам второго столбца).

Все свойства пп. 2 и 3 применяются для вычисления определителей. Пусть, например, надо вычислить определитель

$$D = \begin{vmatrix} 1 & 0 & 2 & -1 \\ 3 & 1 & 0 & -1 \\ 2 & 1 & -1 & 0 \\ 0 & 3 & 2 & 1 \end{vmatrix}.$$

Для этого можно с помощью применения свойства 6 п. 2 сделать в каком-либо из рядов все элементы, кроме одного, равными нулю; тогда, раскладывая полученный определитель по элементам этого ряда, мы получим всего лишь одно слагаемое, так как в остальных алгебраические дополнения множатся на нули. Так, если мы хотим в третьей строчке определителя оставить отличный от нуля элемент лишь на втором месте, то надо второй столбец помножить на —2 и прибавить к первому, а затем в полученном определителе второй столбец прибавить к третьему. Получим

$$D = \begin{vmatrix} 1 & 0 & 2 & -1 \\ 1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ -6 & 3 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 2 & -1 \\ 1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 0 \\ -6 & 3 & 5 & 1 \end{vmatrix}$$

(обычно эти две операции выполняются за один шаг). Разлагая по элементам третьей строки, получим

$$D = -1\cdot\begin{vmatrix} 1 & 2 & -1 \\ 1 & 1 & -1 \\ -6 & 5 & 1 \end{vmatrix}.$$

Здесь можно, например, вычесть вторую строку из первой; получим

$$D = -\begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & -1 \\ -6 & 5 & 1 \end{vmatrix}.$$

Если теперь произвести разложение по элементам первой строки, получим окончательно

$$D = -(-1)\begin{vmatrix} 1 & -1 \\ -6 & 1 \end{vmatrix} = 1\cdot 1 - (-1)(-6) = -5.$$

При вычислении определителей (особенно, высокого порядка), элементы которых являются приближенными числами, применяется аналогичный *метод главного элемента*; покажем этот метод на примере. Пусть надо вычислить определитель

$$D=\begin{vmatrix} -1,37 & 2,15 & 0,76 \\ 2,31 & -1,78 & -4,32 \\ -0,86 & 2,13 & 3,15 \end{vmatrix}.$$

Численное значение *главного*, т. е. наибольшего по абсолютной величине *элемента* (— 4,32), вынесем из строчки или столбца, в которых стоит этот элемент (например, из второй строчки):

$$D=-4,32\begin{vmatrix} -1,37 & 2,15 & 0,76 \\ \boxed{-0,535 \;\; 0,412 \;\; 1} \\ -0,86 & 2,13 & 3,15 \end{vmatrix}.$$

Затем умножаем третий столбец на 0,535 (на —0,412) и прибавляем к первому (ко второму):

$$D=-4,32\begin{vmatrix} -0,963 & 1,837 & 0,76 \\ 0 & 0 & 1 \\ 0,825 & 0,832 & 3,15 \end{vmatrix}=4,32\begin{vmatrix} -0,963 & 1,837 \\ \\ 0,825 & 0,832 \end{vmatrix}.$$

Повторим этот прием еще раз:

$$D=4,32\cdot 1,837\begin{vmatrix} -0,524 & 1 \\ 0,825 & 0,832 \end{vmatrix}=4,32\cdot 1,837\begin{vmatrix} 0 & 1 \\ 1,261 & 0,832 \end{vmatrix}=$$
$$=-4,32\cdot 1,837\cdot 1,261=-10,0.$$

Основы теории определителей заложены в 1750 г. швейцарским математиком Г. Кра́мером (1704—1752).

§ 2. Система линейных алгебраических уравнений

4. Формулы для решения. Мы ограничимся случаем, когда число уравнений равно числу неизвестных, и будем считать, как и раньше, что рассматривается система трех уравнений — система (5). Если, например, мы хотим найти y, то надо умножить первое уравнение (5) на алгебраическое дополнение B_1 элемента b_1 в определителе (7), второе уравнение умножить на B_2, третье — на B_3, после чего полученные результаты почленно сложить. Получим

$$(a_1B_1+a_2B_2+a_3B_3)\,x+(b_1B_1+b_2B_2+b_3B_3)\,y+$$
$$+(c_1B_1+c_2B_2+c_3B_3)\,z=d_1B_1+d_2B_2+d_3B_3. \quad (11)$$

Однако первая скобка равна

$$\begin{vmatrix} a_1 & a_1 & c_1 \\ a_2 & a_2 & c_2 \\ a_3 & a_3 & c_3 \end{vmatrix}. \quad (12)$$

В самом деле, если этот определитель разложить по элементам второго столбца, то мы получим сумму произведений элементов a_1, a_2, a_3 на их алгебраические дополнения, которые в определителе (12) как раз таковы, как дополнения элементов b_1, b_2, b_3 в определителе (7), т. е. равны B_1, B_2, B_3. Определитель (12) по свойству 2 п. 2 равен нулю. Аналогично равна нулю третья скобка в формуле (11), а правая часть равна

$$\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}.$$

Вторая же скобка согласно п. 3 равна самому определителю (7), составленному из коэффициентов при неизвестных. Он называется *определителем системы* (5); мы будем для краткости обозначать его буквой D. Итак, из (11) получим

$$Dy = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}. \tag{13}$$

Аналогично найдем

$$Dx = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}; \quad Dz = \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}. \tag{14}$$

Допустим сначала, что $D \neq 0$. Тогда, как мы покажем в конце п. X.7, *рассматриваемая система имеет одно и только одно решение.* Из (13) и (14) получим формулы для решения

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{D}; \quad y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{D}; \quad z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{D},$$

т. е. каждая из неизвестных равна дроби, у которой в знаменателе стоит определитель системы, а в числителе — определитель, полученный из определителя системы подстановкой столбца свободных членов вместо столбца коэффициентов при данной неизвестной.

Рассмотрим, например, систему уравнений

$$\left.\begin{aligned} x - 2z &= 1, \\ 2x + y - z &= 0, \\ x - 2\ \ + z &= -2; \end{aligned}\right\}$$

пусть нужно найти значение z:

$$z=\frac{\begin{vmatrix}1 & 0 & 1\\ 2 & 1 & 0\\ 1 & -2 & -2\end{vmatrix}}{\begin{vmatrix}1 & 0 & -2\\ 2 & 1 & -1\\ 1 & -2 & 1\end{vmatrix}}=\frac{1(-2+0)-0+1(-4-1)}{1(1-2)-0-2(-4-1)}=\frac{-2-5}{-1+10}=-\frac{7}{9}.$$

Можно вычислить и остальные неизвестные; при этом надо будет подсчитывать только числители, так как в знаменателях будет стоять $D=9$.

Для системы (1) двух уравнений с двумя неизвестными приведенное *правило Крамера* даст как раз формулы (4).

5. Численное решение. Если коэффициенты уравнения являются приближенными числами и, особенно, если число уравнений велико, то применение формул п. 4 неудобно. Тогда применяются разнообразные методы численного решения, из которых мы осветим два на примере системы (5), хотя *эти методы пригодны для систем с любым числом уравнений.*

Метод главных элементов Гаусса (К. Гáусс, 1777—1855, крупнейший немецкий математик, внесший фундаментальный вклад также в астрономию и геодезию) аналогичен методу вычисления определителей, описанному в конце п. 3. Если дана система (5) с численными коэффициентами, то мы смотрим, какой из коэффициентов в левых частях будет наибольшим по абсолютному значению. Пусть, например, это будет b_3. Тогда делим третье уравнение на b_3, что даст

$$a_3'x+y+c_3'z=d_3' \tag{15}$$

(штрихом обозначены новые коэффициенты). Умножим уравнение (15) на $-b_1$ (соответственно $-b_2$) и прибавим к первому (второму) уравнению системы. Тогда y исключится и останется система

$$\left.\begin{aligned} a_1'x+c_1'z=d_1',\\ a_2'x+c_2'z=d_2'. \end{aligned}\right\} \tag{16}$$

Теперь мы смотрим, какой из коэффициентов a_1', c_1', a_2', c_2' наибольший по абсолютной величине. Пусть это a_1'; тогда мы делим на a_1' первое из уравнений (16), что даст

$$x+c_1''z=d_1''. \tag{17}$$

Умножив это уравнение на $-a_2'$ и прибавив ко второму уравнению (16), получим уравнение вида

$$c_2''z=d_2'', \tag{18}$$

т. е. и x исключился. Из (18) находим z; подставляя его в (17), находим x; подставляя x и z в 15), находим .

Метод итераций (ср. п. V. 3) в применении к системе (5) выглядит так. Система переписывается в виде

$$\left.\begin{aligned} x&=\alpha_1 x+\beta_1 y+\gamma_1 z+\delta_1,\\ y&=\alpha_2 x+\beta_2 y+\gamma_2 z+\delta_2,\\ z&=\alpha_3 x+\beta_3 y+\gamma_3 z+\delta_3. \end{aligned}\right\} \qquad (19)$$

Затем выбираются некоторые значения $x=x_0$, $y=y_0$, $z=z_0$ в качестве *нулевого приближения*. Эти значения подставляются в правые части системы (19), что дает *первое приближение* x_1, y_1, z_1; подстановкой его в правые части (19) находим *второе приближение* и т. д. Вообще $(n+1)$-е приближение выражается через n-е по формулам

$$\begin{aligned} x_{n+1}&=\alpha_1 x_n+\beta_1 y_n+\gamma_1 z_n+\delta_1,\\ y_{n+1}&=\alpha_2 x_n+\beta_2 y_n+\gamma_2 z_n+\delta_2,\\ z_{n+1}&=\alpha_3 x_n+\beta_3 y_n+\gamma_3 z_n+\delta_3. \end{aligned} \qquad (20)$$

Если процесс сходится, т. е. при $n\to\infty$ последовательные приближения имеют предел, $x_n\to\bar{x}$; $y_n\to\bar{y}$; $z_n\to\bar{z}$, то, переходя в формулах (20) к пределу при $n\to\infty$, видим, что $x=\bar{x}$, $y=\bar{y}$, $z=\bar{z}$ образуют решение системы (19).

Подобно примерам п. V. 3, *для системы* (19) *метод итераций сходится тем быстрее, чем меньше абсолютные значения коэффициентов при неизвестных в правых частях системы.* Более точные признаки сходимости будут даны в п. XVII. 18.

Система (5) особенно просто разрешима, если она имеет *треугольную форму:*

$$\left.\begin{aligned} a_1 x\phantom{{}+b_2 y+c_3 z}&=d_1,\\ a_2 x+b_2 y\phantom{{}+c_3 z}&=d_2,\\ a_3 x+b_3 y+c_3 z&=d_3. \end{aligned}\right\}$$

В этом случае из первого уравнения системы сразу находим x; подставляя его во второе уравнение, находим y, а затем, подставляя x и y в третье уравнение, находим z. Описанный выше метод Гаусса — это, по существу, метод приведения общей системы (5) к треугольной форме (15), (17), (18).

Различные полезные указания по поводу численного решения систем линейных алгебраических уравнений см. в [3], [17], [47].

6. Случай $D=0$. *Если определитель системы равен нулю, то, как будет показано в п. X. 7, может быть один из двух случаев:*

1. *Система несовместна* (противоречива), т. е. не имеет ни одного решения. Например, такой будет система

$$\left.\begin{aligned} x+2y-z&=1,\\ 2x-y&=3,\\ 3x+y-z&=5, \end{aligned}\right\}$$

так как, складывая первые два уравнения, мы вступаем в противоречие с третьим.

2. *Система имеет бесконечное количество решений*; при этом она обязательно *зависима*, т. е. одно из уравнений является следствием остальных. Например, такой будет система

$$\left.\begin{aligned} x+2y-z&=1, \\ 2x-y&=3, \\ 3x+y-z&=4. \end{aligned}\right\} \qquad (21)$$

Здесь третье уравнение вытекает из первых двух, так как получается в результате их сложения. Поэтому третье уравнение можно не принимать во внимание. Чтобы найти *общее решение системы*, т. е. *совокупность всех решений*, перепишем первые два уравнения в виде

$$\left.\begin{aligned} x+2y&=1+z, \\ 2x-y&=3, \end{aligned}\right\}$$

откуда легко найдем $x=\frac{7+z}{5}$, $y=\frac{-1+2z}{5}$, $z=z$. Здесь z остается произвольным; придавая z всевозможные значения, получим совокупность всех решений; например, при $z=0$ получим решение $x=\frac{7}{5}$, $y=-\frac{1}{5}$, $z=0$; при $z=3$ — решение $x=2$; $y=1$; $z=3$ и т. д. (это — «*частные решения*» системы (21)).

Приведенные примеры являются типичными. Оказывается, что всегда, если определитель системы равен нулю, то между левыми частями системы имеется одно или несколько соотношений. Если такие же соотношения справедливы для правых частей, то система имеет бесконечное количество решений; в противном случае нет ни одного решения.

Все эти случаи можно наглядно показать на системе (1). Как известно из п. II. 9, каждое из уравнений (1) определяет на плоскости x, y прямую линию, так что ищется точка пересечения двух прямых. Условие $D\neq 0$ можно записать в виде $\frac{a_1}{a_2}\neq\frac{b_1}{b_2}$; легко проверить геометрический смысл этого условия: указанные прямые не параллельны, а потому имеют как раз одну точку пересечения. Если $D=0$, то прямые параллельны. Тогда может быть два подслучая: если $\frac{a_1}{a_2}=\frac{b_1}{b_2}\neq\frac{d_1}{d_2}$, то прямые не имеют ни одной общей точки, т. е. система (1) противоречива; если $\frac{a_1}{a_2}=\frac{b_1}{b_2}=\frac{d_1}{d_2}$, то прямые совпадают, уравнения (1) равносильны, т. е. имеется бесконечное количество решений (целая «*прямая решений*»).

Такие неприятности при $D=0$ приводят к практическим затруднениям и в случае, если определитель системы отличен от нуля, но очень мал, так как тогда решение получается с очень невысокой точностью. Поэтому систем с таким определителем следует избегать.

Важным частным случаем является система n линейных *однородных* (т. е. без свободных членов) уравнений с n неизвестными. Например, в случае $n=3$ система имеет вид

$$\left.\begin{aligned} a_1x+b_1y+c_1z&=0;\\ a_2x+b_2y+c_2z&=0;\\ a_3x+b_3y+c_3z&=0. \end{aligned}\right\} \tag{22}$$

Такая система, конечно, имеет *нулевое* (*тривиальное*, т. е. неинтересное) решение $x=y=z=0$. Часто бывает важно выяснить, имеются ли другие, *ненулевые* решения. На основании предыдущего легко дать ответ на этот вопрос. Если определитель системы $D\neq 0$, то в силу п. 4 имеется только одно решение, а значит, ненулевых решений нет. Если же $D=0$, то в силу начала этого пункта система имеет бесконечное число ненулевых решений, так как несовместной в данном случае она быть не может; эти решения находятся так же, как в примере (21).

Так, отбрасывая в системе (22) для определенности третье уравнение и задавая произвольно значение $z=t$, мы приходим к системе уравнений

$$\left.\begin{aligned} a_1x+b_1y+c_1z&=0,\\ a_2x+b_2y+c_2z&=0,\\ z&=t. \end{aligned}\right\}$$

Решая ее по правилам п. 4, получим, если ее определитель $\Delta\neq 0$,

$$x=\frac{\begin{vmatrix} 0 & b_1 & c_1\\ 0 & b_2 & c_2\\ t & 0 & 1 \end{vmatrix}}{\Delta}=\frac{t\begin{vmatrix} b_1 & c_1\\ b_2 & c_2 \end{vmatrix}}{\Delta}=CA_3,$$

где обозначено $\frac{t}{\Delta}=C$ (это — произвольная постоянная). Аналогично получаем $y=CB_3$, $z=CC_3$. Если бы мы взамен третьего уравнения изъяли первое, то подобным же образом получили бы общее решение системы (22) в виде $x=CA_1$, $y=CB_1$, $z=CC_1$.

Иногда приходится рассматривать системы линейных алгебраических уравнений, в которых число уравнений не равно числу неизвестных. Общая теория таких систем указана в п. XI. 5.

ВЕКТОРЫ

§ 1. Линейные действия над векторами

1. Скалярные и векторные величины. *Скаля́рные* и *векторные величины* различаются следующим образом: *первые полностью характеризуются своим численным значением в выбранной системе единиц* (например, температура, работа, плотность и т. п.), *тогда как вторые, кроме численного значения, обладают также направлением в пространстве* (например, сила, скорость и т. п.). Все величины, которые мы рассматривали до сих пор, были скалярными, так что специальное слово «скаляр» можно было не применять; но когда наряду со скалярными величинами рассматриваются векторные, то надо всегда ясно представлять, о величинах какого рода идет речь. Векторную величину (или, что то же, *вектор*) можно изобразить отрезком в пространстве, если условиться о единице масштаба: скажем, при изображении сил принять, что сила в 1 *н* изображается отрезком в 2 *см* и т. п. (см. по этому поводу п. I. 4). При этом такой отрезок будет *ориентированным*, т. е. у него должны быть указаны начало и конец; эта ориентация обозначается стрелкой. Если отвлечься от направления векторной величины, получится модуль этой величины. Таким образом, *модуль вектора — это скаляр*; он имеет размерность рассматриваемой величины и всегда положителен (за исключением нуль-вектора, п. 3). Иногда, имея в виду модуль, говорят о *длине* вектора, например вектора силы; при этом подразумевается, что от длины изображающего отрезка мы на основе выбранного масштаба переходим к величине силы, так что длина вектора силы имеет размерность силы. В математике векторы обычно считаются безразмерными; модуль такого вектора является безразмерным числом. Принято обозначать векторы полужирными латинскими буквами или стрелкой (рис. 139). Модуль вектора обозначается той же буквой, но набранной обычным шрифтом, или снабженной вертикальными черточками (*знаком модуля*): $AB = a = |\mathbf{a}|$.

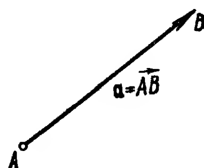


Рис. 139.

Итак, задать вектор — это значит задать его модуль и направление в пространстве. В соответствии с этим каждый вектор можно, не меняя его, перенести *поступательно*, т. е. параллельно самому

себе в любое место; таким образом, начало вектора (*точка* его *приложения*) может находиться где угодно. Другими словами, *два вектора считаются равными, если они имеют одинаковый модуль, параллельны и направлены в одну и ту же сторону* (рис. 140).

Иногда свобода перемещения вектора ограничивается. Например, кроме вектора может быть задана точка его приложения; тогда вектор называется *связанным* (таким является, например, радиус-вектор, см. п. 9). Кроме вектора может быть задана прямая, на которой он должен быть расположен; тогда вектор называется *скользящим*: таким является, например, вектор угловой скорости при вращательном движении, который располагается по оси вращения. Если же перемещение вектора никак не ограничено, то он называется *свободным*.

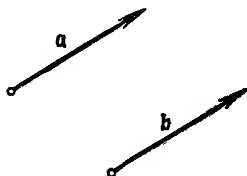


Рис. 140.
$\mathbf{a}=\mathbf{b}$.

2. Сложение векторов. *К линейным действиям* над векторами относятся их *сложение* (и связанное с ним *вычитание*) и *умножение* вектора *на скаляр*. Величина может называться векторной, только если эти действия осуществляются по правилам, описанным в пп. 2—4.

Сложение двух векторов определяется по тому же *правилу параллелограмма*, по которому, как известно из механики, складываются силы и скорости. Если, например, надо сложить два вектора $\mathbf{a}$ и $\mathbf{b}$, то их относят к общему началу, затем на них строят параллелограмм (рис. 141); диагональ этого параллелограмма, выходящая из

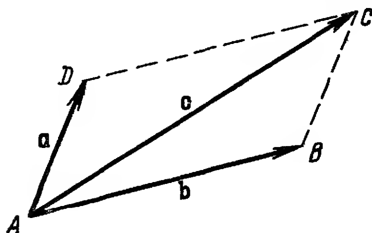


Рис. 141.
$\mathbf{c}=\mathbf{a}+\mathbf{b}$.

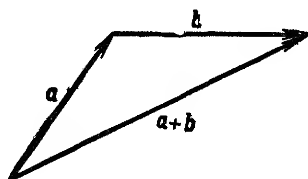


Рис. 142.

того же начала, и даст, по определению, сумму $\mathbf{a}+\mathbf{b}$. Из этого построения сразу видно, что $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$, т. е. *при сложении векторов справедлив переместительный закон.*

Так как противоположные стороны параллелограмма параллельны и равны, то на рис. 141 вектор $\overrightarrow{DC}$ тоже равен $\mathbf{b}$. Отсюда вытекает еще одно правило сложения векторов: к концу первого вектора надо приставить начало второго; тогда *суммой будет служить замыкающий вектор*, т. е. идущий от начала первого в конец второго (рис. 142). Если теперь к этой сумме требуется прибавить некото-

рый третий вектор, то его надо приставить к концу второго вектора и взять замыкающий и т. д. Общее правило сложения любого числа векторов проиллюстрировано на рис. 143. Из рис. 144 вытекает, что справедлив *сочетательный закон* $\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$, который вместе с переместительным законом показывает, что при сложении

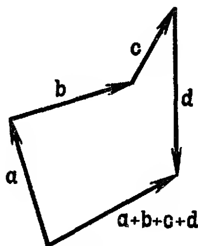


Рис. 143.

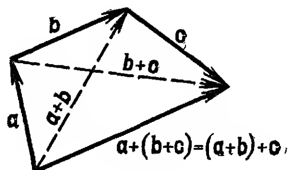


Рис. 144.

любого числа слагаемых их порядок и расстановка скобок несущественны, например,

$$(\mathbf{a}+\mathbf{b})+(\mathbf{c}+\mathbf{d})=[(\mathbf{b}+\mathbf{d})+\mathbf{c}]+\mathbf{a}=[\mathbf{c}+(\mathbf{a}+\mathbf{d})]+\mathbf{b} \text{ и т. п.}$$

Подчеркнем, что *сложение векторов различной размерности*, а также *сложение вектора и скаляра невозможны*; кроме того, в нашем курсе *векторы не сравниваются друг с другом* — не будет положительных и отрицательных векторов, неравенств вида $\mathbf{a}>\mathbf{b}$ и т. д. Конечно, модули (длины) векторов сравнивать друг с другом можно, но не следует удивляться, если модуль суммы векторов окажется меньшим, например, чем модули каждого из слагаемых: ведь векторы складываются не как числа, а как силы, и может оказаться, что равнодействующая нескольких сил окажется меньше, чем каждая из этих сил.

Заметим в заключение следствие из рис. 143:

$$|\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}|\leqslant|\mathbf{a}|+|\mathbf{b}|+|\mathbf{c}|+|\mathbf{d}|;$$

при этом равенство получится, если все слагаемые векторы одинаково направлены; тогда, если их приставить один к другому, они будут продолжать друг друга по прямой.

3. Нуль-вектор и вычитание векторов. *Вектор, конец которого совпадает с его началом, называется нуль-вектором* (нулевым вектором), а иногда просто нулем. Его модуль равен нулю (у всех остальных векторов модуль положителен), а направление не определено: ему можно приписать любое направление, т. е. можно считать его параллельным любому вектору. Он обозначается $\mathbf{0}$ и играет при сложении векторов такую же роль, как число 0 при сложении чисел, так как очевидно, что $\mathbf{a}+\mathbf{0}=\mathbf{a}$.

Если дан вектор $\mathbf{a}=\overrightarrow{AB}$, то вектор $\overrightarrow{BA}$ называется *противоположным* вектором к вектору $\mathbf{a}$ и обозначается $-\mathbf{a}$ (рис. 145). Очевидно, что $\mathbf{a}+(-\mathbf{a})=\mathbf{0}$.

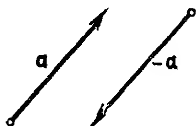


Рис. 145.

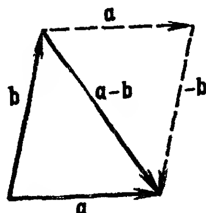


Рис. 146.

Вычесть какой-либо вектор — это значит прибавить противоположный. Отсюда следует, что

$$\mathbf{b}+(\mathbf{a}-\mathbf{b})=\mathbf{b}+[\mathbf{a}+(-\mathbf{b})]=\mathbf{a}+[\mathbf{b}+(-\mathbf{b})]=\mathbf{a}+\mathbf{0}=\mathbf{a},$$

что соответствует обычному определению разности. Геометрическое правило образования разности показано на рис. 146.

4. Умножение вектора на скаляр. Произведение $\lambda\mathbf{a}=\mathbf{a}\lambda$ вектора $\mathbf{a}$ на безразмерный скаляр (число) λ определяется следующим образом: если $\lambda>0$, то это — вектор, получающийся из $\mathbf{a}$ растяжением в λ раз без изменения направления; если же $\lambda<0$, то надо $\mathbf{a}$ растянуть в $|\lambda|$ раз и, кроме того, изменить направление на

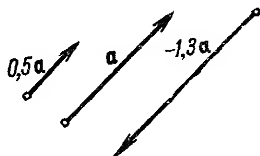


Рис. 147.

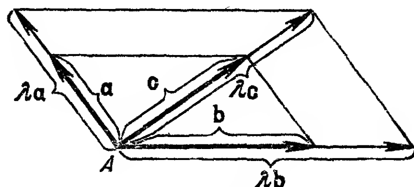


Рис. 148.
$\mathbf{c}=\mathbf{a}+\mathbf{b}$, $\lambda\mathbf{c}=\lambda\mathbf{a}+\lambda\mathbf{b}$.

противоположное (рис. 147). Далее, $\frac{\mathbf{a}}{\lambda}=\left(\frac{1}{\lambda}\right)\mathbf{a}$. Из этих определений вытекают следующие простые свойства:

1. $(-1)\,\mathbf{a}=-\mathbf{a}$;
2. $0\mathbf{a}=\mathbf{0}$;
3. $\lambda\mathbf{0}=\mathbf{0}$;
4. $(\lambda+\mu)\,\mathbf{a}=\lambda\mathbf{a}+\mu\mathbf{a}$;
5. $\lambda(\mathbf{a}+\mathbf{b})=\lambda\mathbf{a}+\lambda\mathbf{b}$;
6. $\lambda(\mu\mathbf{a})=(\lambda\mu)\,\mathbf{a}$;
7. $\lambda\frac{\mathbf{a}}{\lambda}=\mathbf{a}$;
8. $n\mathbf{a}=\underbrace{\mathbf{a}+\mathbf{a}+\ldots+\mathbf{a}}_{n\text{ раз}}$ (если n целое >0).

Все эти свойства дают возможность при линейных действиях с векторами производить преобразования так же, как с числами. Доказательства этих свойств почти очевидны: например, на рис. 148 показано доказательство свойства 5 $[\lambda\mathbf{a}+\lambda\mathbf{b}=\lambda\mathbf{c}=\lambda(\mathbf{a}+\mathbf{b})]$ при $\lambda>0$.

Если скаляр λ размерный, то произведение $\lambda\mathbf{a}$ можно определить как вектор, модуль которого равен $|\lambda||\mathbf{a}|$, причем $\lambda\mathbf{a}$ параллелен $\mathbf{a}$ и направлен в ту же сторону, что $\mathbf{a}$, если $\lambda>0$, и в противоположную, если $\lambda<0$. Все перечисленные свойства остаются справедливыми и в этом случае. Впредь мы для простоты, если не будет особой оговорки, *будем считать все векторы и скаляры безразмерными.*

5. Линейная комбинация векторов. Пусть дано несколько векторов, например три вектора $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. Тогда всякий вектор, имеющий вид $\mathbf{d}=\lambda\mathbf{a}+\mu\mathbf{b}+\nu\mathbf{c}$, где λ, μ, ν — некоторые скаляры, называется *линейной комбинацией* векторов $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$; говорят также, что $\mathbf{d}$ *линейно выражается* через $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, т. е. получается из них с помощью линейных действий. Например, такими линейными комбинациями будут

$$\mathbf{a}+2\mathbf{b}-3\mathbf{c};\quad \frac{\mathbf{a}+\mathbf{c}}{2}\left(=\frac{1}{2}\mathbf{a}+0\mathbf{b}+\frac{1}{2}\mathbf{c}\right);\quad \mathbf{0}\,(=0\mathbf{a}+0\mathbf{b}+0\mathbf{c})\ \text{и т. д.}$$

Заданные векторы называются линейно зависимыми, если какой-либо из этих векторов линейно выражается через остальные; в противном случае эти векторы называются *линейно независимыми* (между собой).

Два вектора линейно зависимы тогда и только тогда, если они параллельны друг другу. Действительно, из определения п. 4 вытекает, что если $\mathbf{b}=\lambda\mathbf{a}$, то $\mathbf{b}\parallel\mathbf{a}$. Наоборот, если два вектора параллельны, то один из них всегда можно растянуть во столько раз, чтобы получить второй (если векторы направлены противоположно друг другу, то коэффициент растяжения будет отрицательным).

Три вектора линейно зависимы тогда и только тогда, когда они параллельны одной плоскости. Действительно, пусть $\mathbf{c}=\lambda\mathbf{a}+\mu\mathbf{b}$; отнесем все три вектора $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ к одному началу и проведем через векторы $\mathbf{a}$ и $\mathbf{b}$ плоскость (P). Тогда векторы $\lambda\mathbf{a}$ и $\mu\mathbf{b}$ будут лежать в плоскости (P), а потому и их сумма, т. е. $\mathbf{c}$, будет лежать в той же плоскости. Значит, и в исходном положении векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ были параллельны плоскости (P). Наоборот, пусть дано, что векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ параллельны одной плоскости; если тогда их отнести к одному началу, то они все будут лежать в одной плоскости (P). Если $\mathbf{a}\nparallel\mathbf{b}$, то представление $\mathbf{c}$ в виде линейной комбинации $\mathbf{a}$ и $\mathbf{b}$ показано на рис. 149; оно называется *разложением вектора в плоскости по двум*

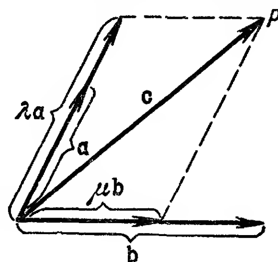


Рис. 149.
$\mathbf{c}=\lambda\mathbf{a}+\mu\mathbf{b}$.

непараллельным векторам. Если же $\mathbf{a}\|\mathbf{b}$, то согласно предыдущему абзацу уже из двух векторов **a**, **b** один (например, **a**) линейно выражается через другой (**b**), т. е. **a** выражается через **b** и **c**.

Четыре или более векторов всегда линейно зависимы. Действительно, пусть даны четыре вектора **a**, **b**, **c**, **d**; отнесем их к одному началу. Если после этого векторы **a**, **b**, **c** окажутся лежащими в одной плоскости, то по предыдущему абзацу один из них линейно выражается через остальные и т. д. (как в конце предыдущего абзаца). Пусть теперь **a**, **b**, **c** не лежат в одной плоскости (рис. 150). Тогда проводим через точку D (конец вектора **d**) прямую, параллельную вектору **c** до пересечения с плоскостью векторов **a** и **b** в точке C; затем через C проводим прямую, параллельную

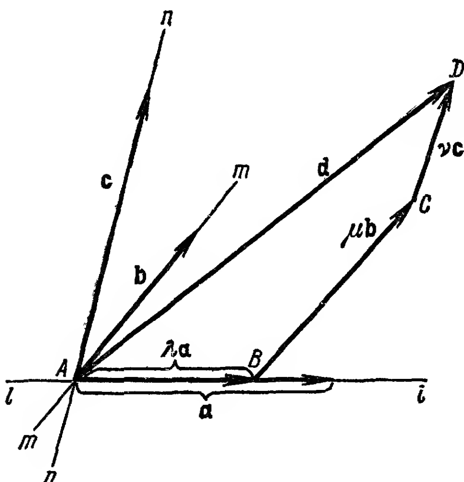


Рис 150.
$\mathbf{d}=\lambda\mathbf{a}+\mu\mathbf{b}+\nu\mathbf{c}$

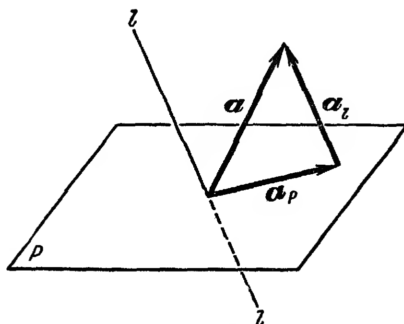


Рис. 151.
$\mathbf{a}=\mathbf{a}_p+\mathbf{a}_l$

вектору **b**, до пересечения с прямой, на которой лежит вектор **a**, в точке B. Тогда

$$\mathbf{d}=\overrightarrow{AD}=\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CD}=\lambda\mathbf{a}+\mu\mathbf{b}+\nu\mathbf{c}. \qquad (1)$$

Это представление называется *разложением вектора по трем векторам, не параллельным одной плоскости*; оно же называется *разложением вектора по трем осям* (на рис. 150 по осям ll, mm и nn). Такое разложение часто применяется в теоретической механике и в других дисциплинах при разложении силы по трем направлениям и т. п. Каждое из слагаемых $\lambda\mathbf{a}$, $\mu\mathbf{b}$ и $\nu\mathbf{c}$ называется *составляющей* или, что то же, *компонéнтой* вектора **d** по (вдоль) соответствующей оси. Компонента вдоль каждой оси определяется не только направлением этой оси, но и направлениями остальных осей. Однако от ориентации осей, т. е. от указания на них положительного направления компоненты не зависят.

Иногда применяется разложение вектора по оси и не параллельной ей плоскости. Это разложение показано на рис. 151.

Разложение (1) возможно осуществить лишь единственным образом. В самом деле, если бы наряду с (1) существовало другое разложение $\mathbf{d}=\lambda_1\mathbf{a}+\mu_1\mathbf{b}+\nu_1\mathbf{c}$, то, приравнивая правые части, мы получим $(\lambda-\lambda_1)\mathbf{a}+(\mu-\mu_1)\mathbf{b}+(\nu-\nu_1)\mathbf{c}=\mathbf{0}$, откуда вытекает линейная зависимость векторов **a**, **b**, **c** (почему?). Если дано более четырех векторов, то уже из первых четырех векторов один линейно выражается через все остальные, т. е. векторы линейно зависимы.

Совокупность линейно независимых векторов, по которым производится разложение остальных векторов, называется *базисом*. Из сказанного выше следует, что *в плоскости базисом могут служить любые два непараллельных вектора, а в пространстве — любые три вектора, не параллельных одной плоскости*. Если **a**, **b**, **c** — такой базис, то в формуле разложения (1) набор чисел λ, μ, ν однозначно определяется вектором **d** и, наоборот, однозначно определяет его; эти числа, λ, μ, ν, являются *координатами* вектора **d** в базисе **a**, **b**, **c** (относительно базиса **a**, **b**, **c**).

§ 2. Скалярное произведение векторов

6. Проекция вектора на ось. Пусть даны вектор $\mathbf{a}=\overrightarrow{AB}$ и ось l (рис. 152). *Проекцией вектора* **a** *на ось* l ($\text{пр}_l\,\mathbf{a}$) *называется длина отрезка $A'B'$ между основаниями перпендикуляров, опущенных из точек A и B на ось l; причем эта длина берется со знаком + или — в зависимости от того, пройдет ли отрезок $A'B'$ в направлении оси или в противоположном направлении*. Аналогично определяется *проекция одного вектора на другой*; тогда перпендикуляры опускаются на этот другой вектор или на его продолжение. Таким образом, *проекция вектора есть скаляр*; размерность ее такая же, как размерность проектируемого вектора.

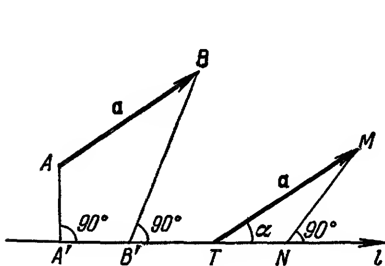


Рис. 152.
Рисунок пространственный!

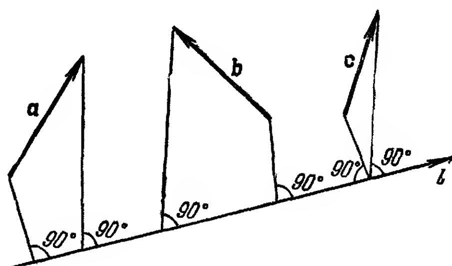


Рис. 153.
$\text{пр}_l\mathbf{a}>0$, $\text{пр}_l\mathbf{b}<0$, $\text{пр}_l\mathbf{c}=0$

Основные свойства проекций следующие:

1. *Знак + или — проекции указывает, идет ли вектор вперед или назад по отношению к оси. Проекция равна нулю* (т. е. A' совпадает с B') *тогда и только тогда, когда вектор перпендикулярен к оси* (рис. 153).

2. *При параллельном переносе вектора его проекция не меняется.*

3. Из рис. 152, из треугольника TNM видно, что

$$\text{пр}_l \mathbf{a} = TM \cos\alpha = a \cos(\widehat{\mathbf{a},\ l}); \qquad (2)$$

здесь скобкой обозначен угол между вектором и осью. В этой формуле знак проекции регулируется знаком косинуса: если угол тупой, то косинус отрицательный, а если острый — косинус положительный, как на рис. 152.

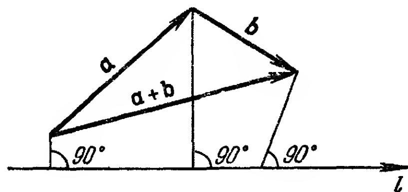


Рис. 154.

4. *Скалярный множитель можно вынести за знак проекции:* $\text{пр}_l(\lambda\mathbf{a}) = \lambda\,\text{пр}_l \mathbf{a}$, так как если вектор увеличить (растянуть) в несколько раз, то и проекция его увеличится во столько же раз.

5. *Проекция суммы равна сумме проекций* (рис. 154): $\text{пр}_l(\mathbf{a}+\mathbf{b}) = \text{пр}_l \mathbf{a} + \text{пр}_l \mathbf{b}$.

7. Скалярное произведение. *Скалярное произведение двух векторов* **a** и **b** по определению *равно произведению модулей этих векторов и косинуса угла между ними.* Обозначается оно точкой (знак скалярного произведения) или круглыми скобками:

$$\mathbf{a}\cdot\mathbf{b} = (\mathbf{a},\ \mathbf{b}) = ab\cos(\widehat{\mathbf{a},\ \mathbf{b}}). \qquad (3)$$

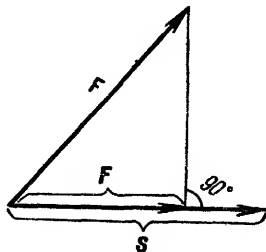


Рис. 155.

Таким образом, *скалярное произведение двух векторов есть скаляр.* Скалярное произведение более чем двух векторов не рассматривается. Имея в виду формулу (2), можно написать также

$$\mathbf{a}\cdot\mathbf{b} = b\,\text{пр}_{\mathbf{b}}\mathbf{a} = a\,\text{пр}_{\mathbf{a}}\mathbf{b}; \qquad (4)$$

скалярное произведение двух векторов равно произведению модуля одного вектора на проекцию на него другого.

Пример. Если точка перемещается на вектор **s** и в процессе этого перемещения на нее действует постоянная сила **F** (рис. 155), то при подсчете работы A, совершенной этой силой, надо учитывать только составляющую $\mathbf{F}'$ силы **F** по направлению перемещения. Таким образом, $A = sF' = s\,\text{пр}_{\mathbf{s}}\mathbf{F} = \mathbf{s}\cdot\mathbf{F}$.

Размерность скалярного произведения равна произведению размерностей его множителей.

Особенно простой вид имеют формулы (3) и (4), если один или оба сомножителя являются *единичными векторами,* т. е. имеют модуль, равный единице. Например, если $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}$ — единичные векторы, то

$$\mathbf{e}_1\cdot\mathbf{e}_2 = \cos(\widehat{\mathbf{e}_1,\ \mathbf{e}_2}); \quad \mathbf{a}\cdot\mathbf{e} = \text{пр}_{\mathbf{e}}\mathbf{a}. \qquad (5)$$

Принято обозначать единичный вектор, идущий по оси l (или вектору $\mathbf{b}$), через $\mathbf{l}^\circ$ (соответственно $\mathbf{b}^\circ$; заметим, что $\mathbf{b}=b\mathbf{b}^\circ$) и называть *ортом* этой оси (этого вектора). Тогда, учитывая формулу (5),

$$\text{пр}_l\,\mathbf{a}=\text{пр}_{\mathbf{l}^\circ}\,\mathbf{a}=\mathbf{a}\cdot\mathbf{l}^\circ,\quad \text{пр}_{\mathbf{b}}\,\mathbf{a}=\mathbf{a}\cdot\mathbf{b}^\circ.$$

8. Свойства скалярного произведения. 1. *Скалярное произведение равно нулю в том и только том случае, если векторы перпендикулярны:*

$$\mathbf{a}\cdot\mathbf{b}=0 \quad \text{равносильно} \quad \mathbf{a}\perp\mathbf{b}.$$

Действительно, это следует из (3) так как $\cos(\widehat{\mathbf{a},\ \mathbf{b}})=0$ означает, что $(\widehat{\mathbf{a},\ \mathbf{b}})=90^\circ$. Конечно, может быть также, что, например, $a=0$, но это значит, что $\mathbf{a}=\mathbf{0}$, а нуль-вектор (см. п. 3) можно считать перпендикулярным любому вектору.

2. $\mathbf{a}\cdot\mathbf{a}=a^2$, так как $(\widehat{\mathbf{a},\ \mathbf{a}})=0^\circ$ и $\cos(\widehat{\mathbf{a},\ \mathbf{a}})=1$. Другими словами, *скалярный квадрат вектора равен квадрату его модуля.*

3. *Скалярное произведение не зависит от порядка сомножителей:* $\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a}$, как это сразу следует из формулы (3).

4. *Скалярный множитель можно выносить за знак скалярного произведения:*

$$(\lambda\mathbf{a})\cdot\mathbf{b}=\mathbf{a}\cdot(\lambda\mathbf{b})=\lambda(\mathbf{a}\cdot\mathbf{b}). \tag{6}$$

Действительно, на основании свойства 4 из п. 6 $\text{пр}_{\mathbf{b}}(\lambda\mathbf{a})=\lambda\text{пр}_{\mathbf{b}}\,\mathbf{a}$. Если обе части умножить на b и воспользоваться формулой (4), получим $\mathbf{b}\cdot(\lambda\mathbf{a})=\lambda(\mathbf{a}\cdot\mathbf{b})$. Отсюда из перестановочности скалярного произведения и вытекает формула (6).

[Выведите формулу (6), исходя непосредственно из определения скалярного произведения.]

5. *Распределительный закон:*

$$(\mathbf{a}+\mathbf{b})\cdot\mathbf{c}=\mathbf{a}\cdot\mathbf{c}+\mathbf{b}\cdot\mathbf{c}.$$

Для доказательства на основе свойства 5 п. 6 пишем $\text{пр}_{\mathbf{c}}(\mathbf{a}+\mathbf{b})=\text{пр}_{\mathbf{c}}\,\mathbf{a}+\text{пр}_{\mathbf{c}}\,\mathbf{b}$, после чего умножим обе части на c.

Эти свойства дают возможность обращаться со скалярным произведением сравнительно просто, например,

$$(\mathbf{a}+2\mathbf{b})\cdot(2\mathbf{a}-3\mathbf{b})=2\mathbf{a}\cdot\mathbf{a}-3\mathbf{a}\cdot\mathbf{b}+4\mathbf{b}\cdot\mathbf{a}-6\mathbf{b}\cdot\mathbf{b}=2a^2+\mathbf{a}\cdot\mathbf{b}-6b^2.$$

§ 3. Декартовы координаты в пространстве

9. Декартовы координаты в пространстве. Пусть дана тройка векторов $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, отнесенных к общему началу в некоторой точке O, принятой за *начало координат*, и не лежащих в одной плоскости. Если выбрано начало координат, то положение произвольной (текущей) точки M в пространстве полностью характеризуется вектором $\mathbf{r}=\overrightarrow{OM}$,

называемым *радиусом-вектором* точки M (рис. 156). Как было доказано в п. 5, можно принять векторы **a**, **b**, **c** за базис и представить $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}$; таким образом, положение точки M характеризуется набором чисел λ, μ, ν, называемых *аффинными координатами* точки M. В соответствии с концом п. 5 можно сказать, что аффинные координаты точки — это координаты ее радиуса-вектора. Таким образом, как и на плоскости (см. п. II.1), каждая точка имеет определенные координаты и, наоборот, по заданным координатам всегда можно построить точку, однако *в пространстве точка имеет три координаты.*

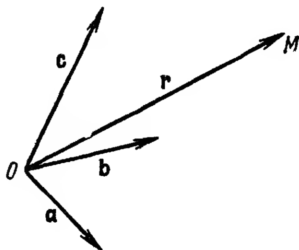


Рис. 156.

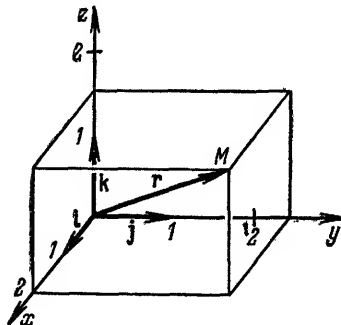


Рис. 157.

Если векторы, принятые за основу системы координат, все имеют единичную длину и взаимно перпендикулярны, то система координат называется декартовой. В этом случае основные векторы (*декартов базис*) принято обозначать буквами **i**, **j**, **k**. Сами декартовы координаты принято обозначать буквами x, y, z; таким образом (рис. 157),

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}. \tag{7}$$

Так, на рис. 157 показана точка M с координатами $x = 2$, $y = 2{,}4$; $z = 1{,}6$. Знак координат указывает на то, в какой из восьми частей, на которые пространство делится плоскостями xOy, yOz и xOz, содержится рассматриваемая точка.

Аналогично формуле (7) любой вектор **a** в декартовой системе координат можно записать в виде

$$\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}, \tag{8}$$

где a_x, a_y, a_z — проекции вектора **a** на соответствующие оси. Если учесть, что в силу формулы (2) для любого единичного вектора **e** будет $\text{пр}_l\,\mathbf{e} = \cos(\widehat{\mathbf{e}, l})$, то из (8) получаем формулу

$$\mathbf{e} = \cos(\widehat{\mathbf{e}, x})\,\mathbf{i} + \cos(\widehat{\mathbf{e}, y})\,\mathbf{j} + \cos(\widehat{\mathbf{e}, z})\,\mathbf{k}.$$

10. Простые задачи на декартовы координаты. 1. *Действия над векторами, разложенными по декартовым осям,* выполняются по очень

простым формулам. Если (см. формулу (8))

$$\mathbf{a}=a_x\mathbf{i}+a_y\mathbf{j}+a_z\mathbf{k},\quad \mathbf{b}=b_x\mathbf{i}+b_y\mathbf{j}+b_z\mathbf{k},\tag{9}$$

то (см. пп. 2, 4 и 8)

$$\mathbf{a}+\mathbf{b}=(a_x+b_x)\,\mathbf{i}+(a_y+b_y)\,\mathbf{j}+(a_z+b_z)\,\mathbf{k};\tag{10}$$

$$\lambda\mathbf{a}=\lambda a_x\mathbf{i}+\lambda a_y\mathbf{j}+\lambda a_z\mathbf{k};\tag{11}$$

$$\mathbf{a}\cdot\mathbf{b}=a_xb_x+a_yb_y+a_zb_z;\tag{12}$$

$$a^2=\mathbf{a}\cdot\mathbf{a}=a_x^2+a_y^2+a_z^2\tag{13}$$

При выводе двух последних очень важных формул надо заметить, что $\mathbf{i}\cdot\mathbf{i}=\mathbf{j}\cdot\mathbf{j}=\mathbf{k}\cdot\mathbf{k}=1$, а $\mathbf{i}\cdot\mathbf{j}=\mathbf{j}\cdot\mathbf{k}=\mathbf{k}\cdot\mathbf{i}=0$, так как векторы **i**, **j**, **k** единичные и взаимно перпендикулярные. Формула (13) — это теорема Пифагора в пространстве: квадрат диагонали прямоугольного параллелепипеда равен сумме квадратов трех его измерений.

Пусть, например, требуется найти угол между векторами $\mathbf{a}=3\mathbf{i}-2\mathbf{j}+\mathbf{k}$ и $\mathbf{b}=-2\mathbf{i}+\mathbf{j}+4\mathbf{k}$. По формуле (3) имеем

$$\cos(\widehat{\mathbf{a},\ \mathbf{b}})=\frac{\mathbf{a}\cdot\mathbf{b}}{ab}=\frac{\mathbf{a}\cdot\mathbf{b}}{\sqrt{a^2}\sqrt{b^2}}=$$

$$=\frac{3(-2)+(-2)1+1\cdot 4}{\sqrt{3^2+(-2)^2+1^2}\sqrt{(-2)^2+1^2+4^2}}=\frac{-4}{\sqrt{14\cdot 21}}=-0{,}233,$$

откуда $(\widehat{\mathbf{a},\ \mathbf{b}})=103\frac{1}{2}^\circ$ в пределах точности логарифмической линейки.

2. *Условия параллельности и перпендикулярности векторов*, заданных своими разложениями (9). Согласно п. 5 условие $\mathbf{a}\|\mathbf{b}$ равносильно тому, что $\mathbf{b}=\lambda\mathbf{a}$ или в силу (11) $b_x=\lambda a_x$; $b_y=\lambda a_y$; $b_z=\lambda a_z$. Исключая λ, получим требуемое условие:

$$\frac{b_x}{a_x}=\frac{b_y}{a_y}=\frac{b_z}{a_z}\qquad (\text{условие } \mathbf{a}\|\mathbf{b}).$$

Далее, согласно п. 8 (свойство 1) условие $\mathbf{a}\perp\mathbf{b}$ равносильно тому, что $\mathbf{a}\cdot\mathbf{b}=0$. Поэтому в силу (12) получаем условие

$$a_xb_x+a_yb_y+a_zb_z=0\qquad (\text{условие } \mathbf{a}\perp\mathbf{b}).$$

3. *Направляющие косинусы* вектора — это косинусы углов, которые он образует с осями координат. Если вектор **a** задан своим разложением (8), то по формуле (2) $a_x=\text{пр}_x\mathbf{a}=a\cos(\widehat{\mathbf{a},\ x})$, т. е.

$$\cos(\widehat{\mathbf{a},\ x})=\frac{a_x}{a};\quad \cos(\widehat{\mathbf{a},\ y})=\frac{a_y}{a};\quad \cos(\widehat{\mathbf{a},\ z})=\frac{a_z}{a}.$$

Отсюда

$$\cos^2(\widehat{\mathbf{a},\ x})+\cos^2(\widehat{\mathbf{a},\ y})+\cos^2(\widehat{\mathbf{a},\ z})=\frac{a_x^2}{a^2}+\frac{a_y^2}{a^2}+\frac{a_z^2}{a^2}=1.$$

Направляющие косинусы вектора полностью определяют его направление, но ничего не говорят о его длине.

Аналогично определяются *направляющие косинусы* какой-либо *оси*: для этого достаточно взять любой вектор, идущий по оси.

4. *Вектор, соединяющий две точки* $M_1(x_1;\ y_1;\ z_1)$ и $M_2(x_2;\ y_2;\ z_2)$ (рис. 158). Имеем, учитывая формулу (7), $\overrightarrow{OM_1} = \mathbf{r}_1 = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}$; $\overrightarrow{OM_2} = \mathbf{r}_2 = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}$, откуда (см. п. 3)

$$\overrightarrow{M_1M_2} = \mathbf{r}_2 - \mathbf{r}_1 = (x_2 - x_1)\,\mathbf{i} + (y_2 - y_1)\,\mathbf{j} + (z_2 - z_1)\,\mathbf{k}.$$

5. *Расстояние между точками* $M_1(x_1;\ y_1;\ z_1)$ и $M_2(x_2;\ y_2;\ z_2)$ в силу предыдущего свойства равно

$$M_1M_2 = \sqrt{(\overrightarrow{M_1M_2})^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \qquad (14)$$

Эта формула очень похожа на соответствующую формулу на плоскости (формула (II. 1)).

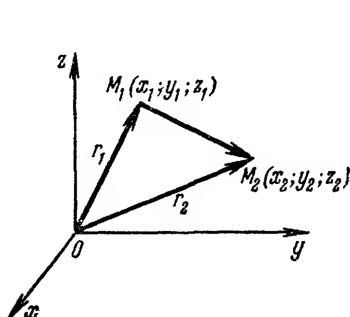


Рис. 158.

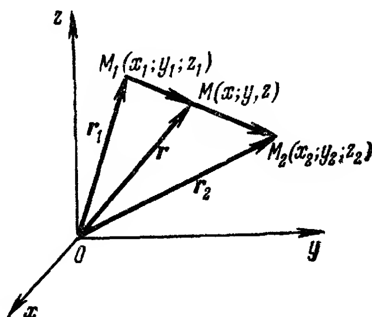


Рис. 159.

6. *Деление отрезка в данном отношении.* Дано (рис. 159)

$$M_1(x_1;\ y_1;\ z_1),\quad M_2(x_2;\ y_2;\ z_2),\quad \frac{M_1M}{MM_2} = \lambda;$$

надо найти точку $M(x;\ y;\ z)$. Имеем

$$\mathbf{r}_1 = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k};\quad \mathbf{r}_2 = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}; \qquad (15)$$

$\overrightarrow{M_1M} = \lambda\overrightarrow{MM_2}$ в силу параллельности этих векторов; но $\overrightarrow{M_1M} = \mathbf{r} - \mathbf{r}_1$; $\overrightarrow{MM_2} = \mathbf{r}_2 - \mathbf{r}$, т. е. $\mathbf{r} - \mathbf{r}_1 = \lambda(\mathbf{r}_2 - \mathbf{r})$; $\mathbf{r} - \mathbf{r}_1 = \lambda\mathbf{r}_2 - \lambda\mathbf{r}$; $\mathbf{r} + \lambda\mathbf{r} = \mathbf{r}_1 + \lambda\mathbf{r}_2$;

$$\mathbf{r} = \frac{\mathbf{r}_1 + \lambda\mathbf{r}_2}{1 + \lambda}. \qquad (16)$$

Приравнивая проекции обеих частей на оси x, y, z (см. формулы (7) и (15)), получим окончательно

$$x = \frac{x_1 + \lambda x_2}{1 + \lambda};\quad y = \frac{y_1 + \lambda y_2}{1 + \lambda};\quad z = \frac{z_1 + \lambda z_2}{1 + \lambda}. \qquad (17)$$

Переход от формулы (16), дающей векторное решение задачи, к формулам (17) называется *проектированием формулы* (16) на оси координат. Ясно, что вообще *всякое векторное равенство вида* $\mathbf{a}=\mathbf{b}$ *в пространстве равносильно трем скалярным*:

$$a_x=b_x, \quad a_y=b_y, \quad a_z=b_z,$$

получающимся в результате проектирования первого на оси координат.

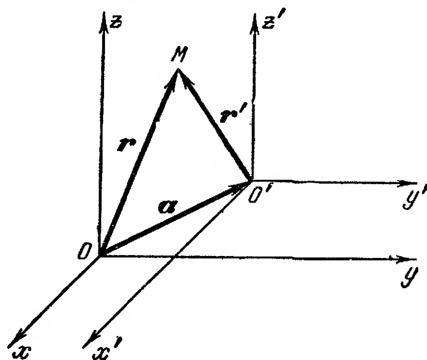


Рис. 160.

7. *Параллельный перенос осей координат.* Пусть оси координат x', y', z' получены из осей x, y, z с помощью параллельного переноса на вектор $\mathbf{a}$ (рис. 160). Тогда, проектируя соотношение $\mathbf{r}=\mathbf{a}+\mathbf{r}'$ между радиусами-векторами любой точки M на оси координат, получаем соотношения между старыми и новыми координатами:

$$x=x'+a_x, \quad y=y'+a_y, \quad z=z'+a_z$$

(см. задачу 3.1 п. II.2).

§ 4. Векторное произведение векторов

11. Ориентация поверхности и вектор площадки. *Поверхность в пространстве называется ориентированной, если указано, какая ее сторона считается наружной, а какая — внутренней.* Как правило, такую ориентацию можно произвести двумя способами (рис. 161); даже для замкнутой поверхности (типа сферы) в некоторых задачах приходится вводить «неестественную» ориентацию, считая, что внутрь (в обычном смысле) обращена наружная сторона этой поверхности.

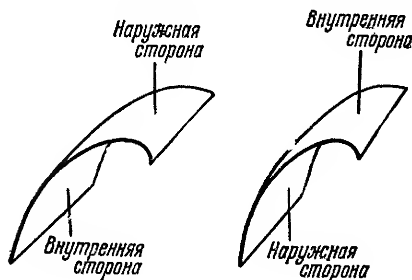


Рис. 161.
Два способа ориентации одной и той же поверхности.

Ориентацию незамкнутой поверхности можно задать также, указав направление обхода ее контура. Таким образом, для незамкнутой поверхности ориентацию можно указывать двумя методами. Чтобы установить связь между ними, надо дополнительно указать, какое

правило применяется — *правого* или *левого винта* (рис. 162). Например, правило правого винта можно сформулировать так: если правый винт (который обычно применяется в технике и обыденной жизни) вращать в направлении обхода контура, то винт должен пойти от внутренней стороны поверхности к наружной. Или, другими словами: если маленький человек обходит по наружной стороне поверхности ее контур в указанном направлении, то пропасть должна быть у него по правую руку, а сама поверхность — по левую.

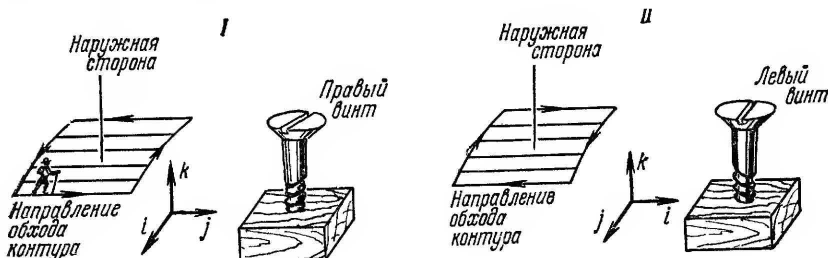


Рис. 162.

Иногда при рассмотрении ориентированного куска плоскости существенными оказываются только его площадь и направление в пространстве, тогда как конкретная форма этого куска — будет ли это круг, или прямоугольник и т. д. — не играет роли. В этом случае кусок (S) плоскости можно изобразить вектором, направленным перпендикулярно к этому куску в направлении от внутренней стороны к наружной (рис. 163), причем модуль этого вектора берется равным площади куска. Такой вектор, называемый *вектором площадки* (S) и обозначаемый $\mathbf{S}$, полностью определяет площадь и направление в пространстве рассматриваемого куска плоскости.

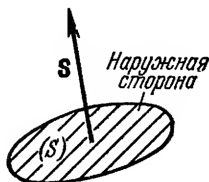


Рис. 163.

Приведем пример применения вектора площадки. Пусть в *однородный* поток газа (т. е. в поток, скорость $\mathbf{v}$ частиц которого во всех точках одинакова) мысленно помещена ориентированная плоская площадка (S), и требуется найти, какой объем газа проходит через (S) в единицу времени изнутри наружу. Так как газ, прошедший за единицу времени, заполняет цилиндр с основанием (S) и высотой, равной $|\mathbf{v}|\cos(\widehat{\mathbf{S}, \mathbf{v}})$ (почему?), то искомый объем равен $|\mathbf{S}|\,|\mathbf{v}|\cos(\widehat{\mathbf{S}, \mathbf{v}}) = \mathbf{S}\cdot\mathbf{v}$.

12. Векторное произведение. Векторное произведение двух векторов $\mathbf{a}$ и $\mathbf{b}$ по определению представляет собой вектор площадки, которая получится, если $\mathbf{a}$ и $\mathbf{b}$ отнести к одному началу, затем построить на этих векторах параллелограмм и обходить его контур, начиная с первого вектора, т. е. с $\mathbf{a}$. Это определение показано на

рис. 164, где принято правило правого винта, которым мы всегда будем пользоваться в нашем курсе, если не оговорено противное.

Введем для дальнейшего понятие «*смысла*» тройки векторов. Пусть даны три вектора **a**, **b**, **c** с общим началом, перечисленных в определенном порядке (первый **a**, второй **b**, третий **c**) и не лежащих в одной плоскости. Такая *тройка векторов* называется *правой*

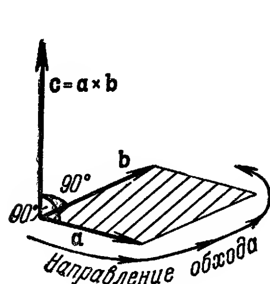


Рис. 164.

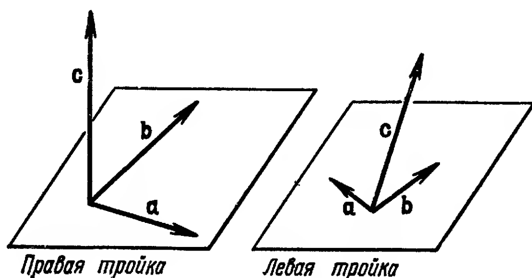


Рис. 165.

или *левой* в зависимости от того, как происходит кратчайший поворот от вектора **a** к вектору **b**, если мы смотрим с конца вектора **c**: *против* часовой стрелки или *по ней* (рис. 165). Происхождение этого названия показано на рис. 166. Отметим, что если переставить номера двух векторов, оставив третий на своем месте, то «смысл» тройки изменится: например, если тройка **a**, **b**, **c** была правой, то тройка **a**, **c**, **b** будет левой (проверьте!). *При циклической* же *перестановке векторов*, т. е. когда второй подставляется вместо первого, третий — вместо второго и первый — вместо третьего (или наоборот), «*смысл*» *тройки не меняется*: если тройка **a**, **b**, **c** была, скажем, правой, то тройка **b**, **c**, **a** также будет правой.

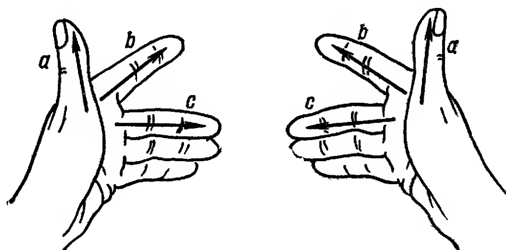


Рис. 166.

Смысл декартовой тройки **i**, **j**, **k** всегда должен соответствовать выбранному правилу винта: тройка берется правой, если выбрано правило правого винта, левой в противном случае; соответственно различаются *правые* (как на рис. 154—156) и *левые декартовы системы координат*. Теперь можно дать новое определение векторного произведения, которое, как видно из рис. 164, равносильно предыдущему.

Векторное произведение двух векторов **a** *и* **b** — *это вектор*, направленный перпендикулярно к обоим векторам, по модулю равный площади параллелограмма, построенного на **a** и **b**, и образующий с этими векторами тройку такого же смысла (т. е. правую или левую), как векторы **i**, **j**, **k**. Обозначается векторное произведение косым крестом $\mathbf{a}\times\mathbf{b}$ или квадратными скобками $[\mathbf{a}, \mathbf{b}]$.

13. Свойства векторного произведения. 1. *Модуль векторного произведения* равен $|\mathbf{a}\times\mathbf{b}| = ab\sin(\widehat{\mathbf{a}, \mathbf{b}})$, так как по такой формуле вычисляется площадь параллелограмма.

Из этой формулы, формулы (3) и свойства 2 п. 8 вытекает следствие

$$(\mathbf{a}\times\mathbf{b})^2+(\mathbf{a}\cdot\mathbf{b})^2=|\mathbf{a}\times\mathbf{b}|^2+(\mathbf{a}\cdot\mathbf{b})^2=a^2b^2\sin^2(\widehat{\mathbf{a}, \mathbf{b}})+a^2b^2\cos^2(\widehat{\mathbf{a}, \mathbf{b}})=a^2b^2;$$

конечно, в левой части первое слагаемое представляет собой скалярный квадрат вектора, а второе — квадрат скаляра.

2. *Векторное произведение равно нуль-вектору в том и только том случае, если векторы параллельны*:

$$\mathbf{a}\times\mathbf{b}=\mathbf{0} \quad \text{равносильно} \quad \mathbf{a}\parallel\mathbf{b},$$

так как параллельность векторов означает вырождение параллелограмма в отрезок, площадь которого равна нулю. В частности, всегда $\mathbf{a}\times\mathbf{a}=\mathbf{0}$.

3. *Векторное произведение антикоммутативно*:

$$\mathbf{b}\times\mathbf{a}=-(\mathbf{a}\times\mathbf{b}).$$

Действительно, при перемене порядка множителей параллелограмм не изменится, но контур его будет проходиться в противоположном направлении и потому вектор площадки заменится на противоположный.

4. *Скалярный множитель можно выносить за знак векторного произведения*:

$$(\lambda\mathbf{a})\times\mathbf{b}=\mathbf{a}\times(\lambda\mathbf{b})=\lambda(\mathbf{a}\times\mathbf{b}),$$

так как если одну из сторон параллелограмма удлинить в λ раз, не меняя ее направления, то и площадь увеличится в λ раз.

5. *Распределительный закон*:

$$(\mathbf{a}+\mathbf{b})\times\mathbf{c}=\mathbf{a}\times\mathbf{c}+\mathbf{b}\times\mathbf{c}, \quad \mathbf{c}\times(\mathbf{a}+\mathbf{b})=\mathbf{c}\times\mathbf{a}+\mathbf{c}\times\mathbf{b}. \tag{18}$$

Для его доказательства отнесем векторы **a**, **b**, **c** к одному началу O, после чего проведем через O плоскость $(P)\perp\mathbf{c}$ (рис. 167). Затем построим на **a** и **b** параллелограмм с диагональю **d** и спроектируем его на плоскость (P), после чего повернем спроектированный параллелограмм в плоскости (P) вокруг оси **c** на 90°, одновременно подобно растянув его в c раз (все это показано на рис. 167). Ясно, что

$$\mathbf{d}=\mathbf{a}+\mathbf{b}, \quad \mathbf{d}'=\mathbf{a}'+\mathbf{b}', \tag{19}$$

где **a**′ и **b**′ — стороны третьего параллелограмма, а **d**′ — его диагональ.

Для удобства то, что происходит с вектором **a**, показано отдельно на рис. 168. Мы видим, что $\mathbf{a}' \perp OK$ и $\mathbf{a}' \perp \mathbf{c}$ (так как $\mathbf{c} \perp (P)$), откуда вектор $\mathbf{a}'$ перпендикулярен ко всей плоскости KOM. Кроме того, $a' = cOK_1$, а последнее произведение как раз равно площади параллелограмма $OKLM$, построенного на векторах **a** и **c** (почему?). Значит, по определению векторного произведения

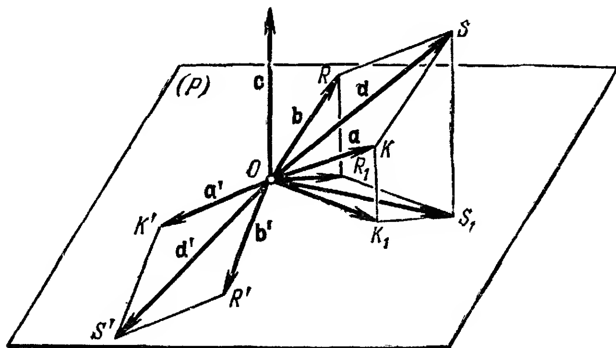


Рис. 167.
$OK_1S_1R_1$ — спроектированный параллелограмм; $OK'S'R'$ — повернутый и растянутый параллелограмм.

$\mathbf{a}' = \mathbf{a} \times \mathbf{c}$. Аналогично $\mathbf{b}' = \mathbf{b} \times \mathbf{c}$; $\mathbf{d}' = \mathbf{d} \times \mathbf{c}$. Из формул (19) получаем $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = \mathbf{d} \times \mathbf{c} = \mathbf{d}' = \mathbf{a}' + \mathbf{b}' = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$, т. е. первую формулу (18). Переставив порядок множителей и поменяв знаки, получим вторую формулу (18).

Эти свойства дают возможность раскрывать скобки в выражениях, включающих векторное произведение, однако при этом надо тщательно следить за порядком множителей.

Приведем пример:

$$(\mathbf{a} + 2\mathbf{b}) \times (2\mathbf{a} - 3\mathbf{b}) = \\ = 2\mathbf{a} \times \mathbf{a} - 3\mathbf{a} \times \mathbf{b} + \\ + 4\mathbf{b} \times \mathbf{a} - 6\mathbf{b} \times \mathbf{b} = -7\mathbf{a} \times \mathbf{b}.$$

Отметим, что *векторное произведение не обладает свойством сочетательности*: $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$ не обязано равняться $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$. Поэтому выражений вида $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$ (без скобок) следует избегать.

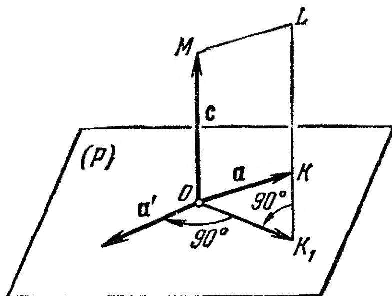


Рис. 168.

6. *Векторное произведение в декартовых проекциях*.

Пусть векторы **a** и **b** даны своими разложениями (9). Тогда, если воспользоваться равенствами, справедливыми независимо от выбора правила винта (проверьте их!),

$$\mathbf{i} \times \mathbf{j} = \mathbf{k};\quad \mathbf{j} \times \mathbf{i} = -\mathbf{k};\quad \mathbf{j} \times \mathbf{k} = \mathbf{i};\quad \mathbf{k} \times \mathbf{j} = -\mathbf{i};\quad \mathbf{k} \times \mathbf{i} = \mathbf{j};$$
$$\mathbf{i} \times \mathbf{k} = -\mathbf{j},$$

получим

$$\mathbf{a}\times\mathbf{b}=(a_x\mathbf{i}+a_y\mathbf{j}+a_z\mathbf{k})\times(b_x\mathbf{i}+b_y\mathbf{j}+b_z\mathbf{k})=$$
$$=a_xb_y\mathbf{k}-a_xb_z\mathbf{j}-a_yb_x\mathbf{k}+a_yb_z\mathbf{i}+a_zb_x\mathbf{j}-a_zb_y\mathbf{i}=$$
$$=\mathbf{i}(a_yb_z-a_zb_y)-\mathbf{j}(a_xb_z-a_zb_x)+\mathbf{k}(a_xb_y-a_yb_x).\quad(20)$$

Полученный результат очень просто запомнить, если записать его в виде определителя [см. формулу (VI.8)]

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_x & a_y & a_z\\ b_x & b_y & b_z\end{vmatrix}.$$

Замечательная формула!

Пусть, например, надо вычислить площадь S параллелограмма, построенного на векторах $\mathbf{a}=3\mathbf{i}-2\mathbf{j}+\mathbf{k}$ и $\mathbf{b}=-2\mathbf{i}+\mathbf{j}+4\mathbf{k}$. Так как $S=|\mathbf{a}\times\mathbf{b}|$, то вычисляем

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 3 & -2 & 1\\ -2 & 1 & 4\end{vmatrix}=\mathbf{i}(-8-1)-\mathbf{j}(12+2)+\mathbf{k}(3-4)=-9\mathbf{i}-14\mathbf{j}-\mathbf{k};$$

$$S=|\mathbf{a}\times\mathbf{b}|=\sqrt{9^2+14^2+1^2}=16{,}7.$$

Результат получается безразмерным, так как и векторы $\mathbf{a}$, $\mathbf{b}$ были безразмерными. Чтобы получилась «настоящая» площадь, надо было бы задавать $\mathbf{a}=3\mathbf{i}$ *см* и т. п.

Отметим еще одну полезную формулу. Пусть на плоскости x, y даны два вектора, $\mathbf{a}=a_x\mathbf{i}+a_y\mathbf{j}$, $\mathbf{b}=b_x\mathbf{i}+b_y\mathbf{j}$, и требуется найти площадь S параллелограмма, построенного на $\mathbf{a}$ и $\mathbf{b}$. Так как

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_x & a_y & 0\\ b_x & b_y & 0\end{vmatrix}=\begin{vmatrix}a_x & a_y\\ b_x & b_y\end{vmatrix}\mathbf{k},$$

то из геометрического смысла векторного произведения вытекает (проверьте!), что

$$\begin{vmatrix}a_x & a_y\\ b_x & b_y\end{vmatrix}=\pm S,\ \text{т. е.}\ S=\left|\begin{vmatrix}a_x & a_y\\ b_x & b_y\end{vmatrix}\right|;\quad(21)$$

здесь берется + или — в зависимости от того, совпадает направление кратчайшего поворота от $\mathbf{a}$ к $\mathbf{b}$ с направлением поворота от $\mathbf{i}$ к $\mathbf{j}$ или нет.

Одним из важных применений векторного произведения является понятие *момента* $\text{мом}_O\mathbf{a}$ какого-либо вектора $\mathbf{a}$, приложенного в определенной точке M, относительно некоторой фиксированной точки O: по определению $\text{мом}_O\mathbf{a}=\overrightarrow{OM}\times\mathbf{a}=\mathbf{r}\times\mathbf{a}$. В физике рассматривают моменты силы, скорости, количества движения и т. д. Если изменить точку приложения вектора $\mathbf{a}$, то его момент, вообще говоря, изменится. Однако если перенести $\mathbf{a}$ *по линии его действия* началом в некоторую точку M', то $\overrightarrow{MM'}=\lambda\mathbf{a}$ и потому, обозначив на минуту буквой $\mathbf{a}'$ вектор $\mathbf{a}$ в новом положении, получим

$$\text{мом}_O\mathbf{a}'=\mathbf{r}'\times\mathbf{a}=(\mathbf{r}+\lambda\mathbf{a})\times\mathbf{a}=\mathbf{r}\times\mathbf{a}=\text{мом}_O\,\mathbf{a}.$$

Итак, момент от такого переноса не меняется, т. е. *в определении момента вектор* $\mathbf{a}$ *можно считать скользящим* (п. 1).

14. Истинные векторы и псевдовекторы. Имеются векторы, которые *зависят от того, принято за основу правило правого или левого винта,* причем если одно правило заменяется на другое, то вектор меняется на противоположный. Такие векторы называются *псевдовекторами* (или *аксиальными,* т. е. *осевыми векторами*) в отличие от *«истинных» векторов,* направление которых не зависит от выбора правила винта. Например, при поступательном движении твердого тела вектор скорости по своему физическому смыслу не зависит от выбора правила винта и потому является истинным вектором. В отличие от этого вектор угловой скорости $\boldsymbol{\omega}$ при вращательном движении твердого тела, который откладывается на оси вращения и по модулю равен численному значению скорости, является псевдовектором, поскольку направление его зависит от выбора правила винта (рис. 169).

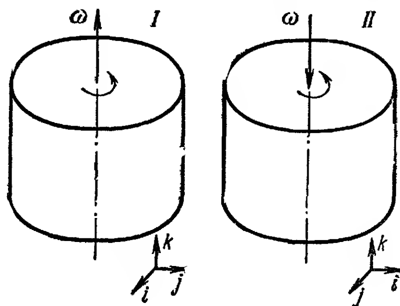


Рис. 169.

Из определения векторного произведения ясно, что векторное произведение двух истинных векторов представляет собой псевдовектор, так как при перемене правила винта бывшая наружная сторона параллелограмма, построенного на векторах **a** и **b**, станет внутренней, и наоборот. Таким образом, например, момент силы (см. конец п. 13)—это псевдовектор. Аналогичным образом векторное произведение истинного вектора на псевдовектор представляет собой истинный вектор, а двух псевдовекторов—также псевдовектор. Так, легко проверить, что истинный вектор **v** линейной скорости любой точки M при вращательном движении связан с псевдовектором $\boldsymbol{\omega}$ формулой $\mathbf{v}=\boldsymbol{\omega}\times\mathbf{r}$ ($\mathbf{r}=\overrightarrow{OM}$), если точка O произвольно выбрана на оси вращения.

Иногда наряду с *«истинными» скалярами* различают также *псевдоскаляры*—скалярные величины, которые при перемене выбора правила винта множатся на -1. Так, легко проверить, что скалярное произведение истинного вектора и псевдовектора является псевдоскаляром.

§ 5. Произведения трех векторов

15. Векторно-скалярное произведение. *Векторно-скалярное произведение трех векторов* **a**, **b**, **c**—это скалярная величина $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$. Ее геометрический смысл виден из рис. 170:

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=\mathbf{d}\cdot\mathbf{c}=d\,\text{пр}_\mathbf{d}\,\mathbf{c}=|\mathbf{a}\times\mathbf{b}|\,\text{пр}_\mathbf{d}\,\mathbf{c}=Sh=V,$$

т. е. получается объем параллелепипеда, построенного на векторах **a**, **b**, **c**. На рис. 170 векторы **a**, **b**, **c** образуют правую тройку и получается объем со знаком $+$. Если бы тройка была левой, то угол между **c** и **d** был бы тупой; в этом случае $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=-V$. (При этом предполагается, что за основу принято правило правого винта.)

Отметим следующие свойства векторно-скалярного произведения.

1. *При циклической перестановке множителей оно не меняется:*

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=(\mathbf{b}\times\mathbf{c})\cdot\mathbf{a}=(\mathbf{c}\times\mathbf{a})\cdot\mathbf{b},$$

так как при такой перестановке не меняются ни параллелепипед (рис. 170), ни «смысл» тройки векторов-множителей (см. п. 12). При перестановке же только двух множителей знак векторно-скалярного произведения меняется: например, $(\mathbf{c}\times\mathbf{b})\cdot\mathbf{a}=-(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$.

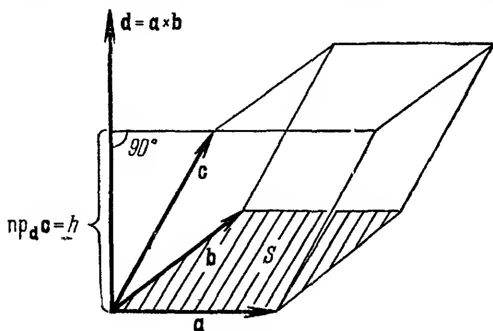


Рис. 170.

2. *Оно равно нулю тогда и только тогда, когда все три вектора параллельны одной плоскости.* Действительно, такая параллельность означает, что параллелепипед вырождается в часть плоскости, т. е. имеет нулевой объем.

3. *Его выражение в декартовых проекциях* таково: в силу формул (20) и (12)

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=[(a_yb_z-a_zb_y)\,\mathbf{i}-(a_xb_z-a_zb_x)\,\mathbf{j}+(a_xb_y-a_yb_x)\,\mathbf{k}]\cdot(c_x\mathbf{i}+c_y\mathbf{j}+c_z\mathbf{k})=$$
$$=(a_yb_z-a_zb_y)\,c_x-(a_xb_z-a_zb_x)c_y+(a_xb_y-a_yb_x)\,c_z$$

или, окончательно,

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=\begin{vmatrix}a_x & a_y & a_z\\ b_x & b_y & b_z\\ c_x & c_y & c_z\end{vmatrix}$$

(для проверки разложите определитель по последней строке, см. п. VI.3).

В частности, привлекая свойство 2, получаем условие, необходимое и достаточное для того, чтобы три вектора **a**, **b**, **c**, заданных своими разложениями по декартовым осям, были параллельны одной плоскости:

$$\begin{vmatrix}a_x & a_y & a_z\\ b_x & b_y & b_z\\ c_x & c_y & c_z\end{vmatrix}=0. \qquad (22)$$

Векторно-скалярное произведение трех истинных векторов (см. п. 14) представляет собой скалярное произведение псевдовектора на истинный вектор, т. е. псевдоскаляр.

16. Векторно-векторное произведение. *Векторно-векторное произведение трех векторов* **a**, **b**, **c** — это вектор $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$. Особенного геометрического смысла оно не имеет, но выражается с помощью формулы, которая иногда

применяется. Для вывода этой формулы представим себе, что мы выбрали оси декартовых координат так, что ось x пошла по вектору $\mathbf{a}$, а ось y лежит в плоскости векторов $\mathbf{a}$ и $\mathbf{b}$ (рис. 171). Тогда вектор $\mathbf{a}$ будет иметь проекцию только на ось x, т. е. $\mathbf{a}=a_x\mathbf{i}$ аналогично $\mathbf{b}=b_x\mathbf{i}+b_y\mathbf{j}$, $\mathbf{c}=c_x\mathbf{i}+c_y\mathbf{j}+c_z\mathbf{k}$. Отсюда

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_x & 0 & 0\\ b_x & b_y & 0\end{vmatrix}=a_xb_y\mathbf{k},$$

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 0 & 0 & a_xb_y\\ c_x & c_y & c_z\end{vmatrix}=-\mathbf{i}a_xb_yc_y+\mathbf{j}a_xb_yc_x=$$

$$=a_xc_x(b_x\mathbf{i}+b_y\mathbf{j})-(b_xc_x+b_yc_y)\,a_x\mathbf{i}$$

(проверьте!). Пользуясь формулой (12), получим окончательно

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=(\mathbf{a}\cdot\mathbf{c})\,\mathbf{b}-(\mathbf{b}\cdot\mathbf{c})\,\mathbf{a}.$$

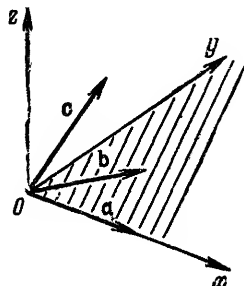


Рис. 171.

Эта формула уже не содержит никаких координатных проекций, а потому не зависит от специального выбора системы координат.

Иногда применяется также следующая формула:

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=-(\mathbf{b}\times\mathbf{c})\times\mathbf{a}=-[(\mathbf{b}\cdot\mathbf{a})\,\mathbf{c}-(\mathbf{c}\cdot\mathbf{a})\,\mathbf{b}]=(\mathbf{a}\cdot\mathbf{c})\,\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}.$$

Отметим в заключение, что векторная алгебра — сравнительно поздняя математическая дисциплина. Она оформилась и развилась во второй половине XIX в. в связи с задачами алгебры, геометрии, механики и физики.

§ 6. Линейные пространства

17. Понятие линейного пространства. Одной из характерных черт векторов является возможность производить над ними линейные действия — сложение и умножение на числа (§ 1). Такие действия возможно производить и над другими объектами, — например, над многочленами или над произвольными функциями, — причем эти действия во всех случаях обладают одинаковыми свойствами. Это дает основание для общего рассмотрения любых *линейных пространств*, т. е. *совокупностей каких-то объектов, над которыми можно, оставаясь в рамках рассматриваемой совокупности, выполнять линейные действия*. Такое общее, абстрактное, рассмотрение приводит к выработке общего взгляда на линейные действия, который позволяет находить важные свойства и в конкретных примерах.

Пусть (R) — *совокупность* (*множество*) некоторых объектов, тогда сами эти объекты называются *элементами* (R); если a — какой-то из этих объектов, то про него можно сказать также, что он *принадлежит* (R), что записывается так: $a\in(R)$; здесь $\in$ — *знак*

принадлежности. Например, если (*Ц*) — множество всех целых чисел, то $3 \in (Ц)$, $-5 \in (Ц)$, $\pi \overline{\in} (Ц)$; здесь $\overline{\in}$ — *знак непринадлежности*.

Переходим к точному определению понятия линейного пространства. Некоторое множество (R) называется линейным пространством, если для любых $\mathbf{x} \in (R)$, $\mathbf{y} \in (R)$ как-то определено понятие суммы $\mathbf{x} + \mathbf{y} \in (R)$, а для любого вещественного числа λ определено $\lambda\mathbf{x} = \mathbf{x}\lambda \in (R)$. Например, пространство R может представлять собой совокупность всех векторов, для которых сложение и умножение на число производятся по правилам § 1. Пространство R может представлять собой совокупность всех комплексных чисел, для которых правила сложения и умножения на вещественные числа известны из курса средней школы. Некоторые дальнейшие примеры будут указаны в п. 18. При этом должны удовлетворяться определенные естественные свойства этих действий, которые в § 1 доказывались, а в общем случае требуются заранее, как *аксиомы линейного пространства*. Эти свойства, в общем, те же, что для векторов; поэтому элементы линейного пространства часто называются (*обобщенными*) *векторами* и обозначаются так же, как обычные векторы.

Мы приведем аксиомы линейного пространства, не заботясь о их *независимости*: другими словами, часть из этих аксиом будет следствием остальных. Сумма должна удовлетворять следующим требованиям:

1. *ассоциативность*, т. е. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ для любых $\mathbf{x}, \mathbf{y}, \mathbf{z} \in (R)$;
2. *коммутативность*, т. е. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$;
3. *существование в* (R) *нулевого элемента*, т. е. элемента, обозначаемого $\mathbf{0}$, для которого $\mathbf{x} + \mathbf{0} = \mathbf{x}$;
4. *существование противоположного элемента* для любого $\mathbf{x} \in (R)$, т. е. элемента $-\mathbf{x} \in (R)$, для которого $(-\mathbf{x}) + \mathbf{x} = \mathbf{0}$.

Легко проверить, что нулевой элемент в пространстве может быть только один и что у каждого элемента может быть только один противоположный. Мы не будем останавливаться на общем доказательстве этих фактов, тем более, что во всех конкретных примерах эти утверждения очевидны.

Умножение элементов на числа должно удовлетворять требованиям:

5. $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$; 6. $1\mathbf{x} = \mathbf{x}$;

7. $(-1)\mathbf{x} = -\mathbf{x}$; 8. $0\mathbf{x} = \mathbf{0}$; 9. $\lambda\mathbf{0} = \mathbf{0}$.

Деление на число вводится по формуле $\frac{\mathbf{x}}{\lambda} = \frac{1}{\lambda}\mathbf{x}$ $(\lambda \neq 0)$.

Наконец, оба линейных действия должны быть связаны друг с другом законами *дистрибутивности*:

10. $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$; 11. $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$.

Все эти свойства дают возможность при линейных действиях в линейном пространстве производить преобразования по обычным, «школьным» правилам. Подробно свойства линейных пространств рассматриваются в курсах линейной алгебры; см., например, [10].

Иногда приходится рассматривать совокупность, в которой определено только сложение, удовлетворяющее аксиомам 1—4. Такая совокупность называется *абелевой группой* по имени одного из наиболее выдающихся математиков XIX в., норвежского ученого Н. Абеля (1802—1829). Кроме того, иногда в линейном пространстве возможно производить умножение не только на вещественные, но и на любые комплексные числа; тогда говорят, что рассматривается *линейное пространство над полем комплексных чисел.* При этом *числовым полем* называется любая совокупность чисел, в которой можно производить четыре арифметических действия, за исключением деления на нуль: так, все действительные числа или все комплексные числа образуют поле, тогда как совокупность целых чисел не образует поля (почему?).

18. Примеры. 1. Простейший пример линейного пространства образует совокупность всех обычных векторов с линейными действиями, описанными в § 1. Совокупность всех векторов, параллельных некоторой плоскости, также образует линейное пространство, которое является *линейным подпространством* первого пространства. Линейным подпространством первого пространства является также совокупность всех векторов, параллельных некоторой прямой; с формальной точки зрения нуль-вектор сам по себе тоже образует линейное пространство, так как для него все аксиомы 1—11 выполнены.

В общем случае некоторое множество (R_1), содержащееся в линейном пространстве (R), называется *линейным подпространством* (R), если (R_1) относительно линейных действий, определенных в (R), само образует линейное пространство. Другими словами, при $\mathbf{x} \in (R_1)$, $\mathbf{y} \in (R_1)$ должно быть $\mathbf{x}+\mathbf{y} \in (R_1)$ и $\lambda\mathbf{x} \in (R_1)$ для любого вещественного λ; проверять аксиомы 1—11 в (R_1) уже не нужно, так как они выполнены во всем (R).

2. Линейное пространство образует совокупность всех многочленов $P(x)$ степени $\leqslant n$, где n — некоторое заданное число. Когда n принимает значения 0, 1, 2 и т. д., каждое следующее пространство содержит предыдущее в качестве линейного подпространства. Все они содержатся в линейном пространстве многочленов всех степеней. Еще более широкое линейное пространство образует совокупность всех функций $f(x)$, определенных на некотором фиксированном интервале. Таким образом, каждую такую функцию можно рассматривать как обобщенный вектор из линейного пространства функций, как говорят, из *функционального пространства.* Такой взгляд на функции характерен для современной математики.

3. Важнейшим примером линейного пространства является совокупность всех (обобщенных) векторов в так называемом *n-мерном вещественном числовом пространстве E_n*, где $n=1, 2, 3, \ldots$ — заданное число.

Рассмотрим, например, случай $n=4$. Точкой в E_4 считается любой набор четырех вещественных чисел $(a_1; a_2; a_3; a_4)$, которые называются *координатами* этой точки; например, $(-3; 0; 2; 1{,}3)$ — это одна точка, $(0; 0; 0; 0)$ — другая точка, называемая *началом координат* в E_4 и т. д. Всё E_4 представляет собой совокупность всех таких точек.

Особенно наглядно числовое пространство E_3. Оно получается, если в обычное пространство ввести аффинную или, в частности, декартову систему координат (§ 3), «приклепать ее намертво», а затем вместо любой точки рассматривать набор ее координат. (Как получить E_2 и E_1?) Преимущество числового подхода состоит в том, что он не ограничен размерностью, пространства E_n рассматриваются совершенно аналогично при всех $n=1, 2, 3, 4, 5, \ldots$

Вернемся к E_4. Уславливаются, что каждая пара точек $A(a_1;\ a_2;\ a_3;\ a_4)$, $B(b_1;\ b_2;\ b_3;\ b_4)$ определяет *обобщенный вектор* $\overrightarrow{AB}$ с началом в A и концом в B; строго говоря, такой вектор — в дальнейшем мы будем опускать слово «обобщенный» — *это и есть* пара точек, взятых в определенном порядке. Так как векторы должны быть свободными (см. п. 1), то уславливаются, что пара точек $A'(a_1+\alpha;\ a_2+\beta;\ a_3+\gamma;\ a_4+\delta)$ и $B'(b_1+\alpha;\ b_2+\beta;\ b_3+\gamma;\ b_4+\delta)$ при любых α, β, γ, δ определяет тот же вектор, что пара A, B, т. е. $\overrightarrow{A'B'}=\overrightarrow{AB}$, вектор $\overrightarrow{A'B'}$ получается из $\overrightarrow{AB}$ в результате параллельного переноса. С помощью такого переноса можно начало любого вектора отнести в любую точку E_4. Так, положив $\alpha=-a_1$, $\beta=-a_2$, $\gamma=-a_3$, $\delta=-a_4$, мы перенесем начало вектора $\overrightarrow{AB}$ в начало координат O, а конец — в точку $M(x_1;\ x_2;\ x_3;\ x_4)$, где $x_1=b_1-a_1, \ldots$ $\ldots, x_4=b_4-a_4$; вектор $\overrightarrow{OM}=\overrightarrow{AB}$ естественно называть *радиусом-вектором* точки M.

Разности между одноименными координатами конца и начала вектора, т. е. $b_1-a_1, \ldots, b_4-a_4$, *называются координатами этого вектора*; они не меняются при его параллельном переносе. Итак, вектор в числовом пространстве, или, как говорят, *числовой вектор*, полностью определяется своими координатами. Зная координаты вектора и координаты его начала, легко найти координаты его конца.

Например, если вектор $\mathbf{x}(2;\ -5;\ 0;\ 1)$ отложить от точки $A(51;\ 0;\ 2;\ -4)$, то его конец будет в точке $B(53;\ -5;\ 2;\ -3)$, а если тот же вектор отложить от точки $O(0;\ 0;\ 0;\ 0)$, то конец будет в точке $M(2;\ -5;\ 0;\ 1)$.

Линейные действия над векторами, заданными своими координатами, определяются аналогично формулам (10) и (11): если даны векторы $\mathbf{x}(x_1;\ x_2;\ x_3;\ x_4)$ и $\mathbf{y}(y_1;\ y_2;\ y_3;\ y_4)$, то $\mathbf{x}+\mathbf{y}$ имеет координаты $(x_1+y_1;\ x_2+y_2;\ x_3+y_3;\ x_4+y_4)$, а $\lambda\mathbf{x}$ — координаты $(\lambda x_1;\ \lambda x_2;\ \lambda x_3;\ \lambda x_4)$. Легко проверить, что все аксиомы линейного пространства выполнены, причем $\mathbf{0}(0;\ 0;\ 0;\ 0)$, $-\mathbf{x}(-x_1;\ -x_2;\ -x_3;\ -x_4)$.

19. Размерность линейного пространства. Пусть рассматривается линейное пространство (R). Понятия *линейной комбинации* и *линейной зависимости* векторов в нем вводятся в точности, как в п. 5. Однако, вообще говоря, четыре вектора в (R) уже не обязаны быть линейно зависимыми. Могут представиться два случая.

1. В (R) можно указать некоторое количество n линейно независимых векторов, но всякие $n+1$ векторов будут линейно зависимыми между собой. Тогда пространство (R) называется *n-мерным*, а любая совокупность из n линейно независимых векторов в нем называется *базисом*. Итак, *размерность линейного пространства — это максимально возможное число линейно независимых векторов в нем*.

2. В (R) можно указать как угодно большое количество линейно независимых векторов. Тогда пространство (R) называется *бесконечномерным*.

Это определение размерности хорошо согласуется с обычным пониманием размерности, так как в силу п. 5 получается, что пространство обычных векторов трехмерно, а пространство векторов, лежащих на плоскости (на прямой), двумерно (соответственно одномерно). Пространство, состоящее из одного только нуль-вектора, формально считается *нульмерным*.

При подсчете размерности конкретных пространств оказывается полезной следующая простая лемма *пусть каждый из векторов* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ *представляет собой линейную комбинацию векторов* $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l$, *причем* $k>l$; *тогда векторы* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ *линейно зависимы между собой*.

Мы не будем давать общего доказательства этой леммы, а поясним ее на частном случае. Пусть, например, векторы $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ являются линейными комбинациями векторов $\mathbf{y}_1, \mathbf{y}_2$, т. е.

$$\mathbf{x}_1=\alpha\mathbf{y}_1+\beta\mathbf{y}_2, \quad \mathbf{x}_2=\gamma\mathbf{y}_1+\delta\mathbf{y}_2, \quad \mathbf{x}_3=\varepsilon\mathbf{y}_1+\zeta\mathbf{y}_2. \tag{23}$$

Если определитель $D=\begin{vmatrix}\alpha & \beta \\ \gamma & \delta\end{vmatrix}\neq 0$, то первые два равенства можно рассмотреть как уравнения относительно $\mathbf{y}_1, \mathbf{y}_2$ и решить их, линейно выразив $\mathbf{y}_1, \mathbf{y}_2$ через $\mathbf{x}_1, \mathbf{x}_2$; подставив полученные выражения в третье равенство (23), получаем линейную зависимость между $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. Если же $D=0$, то правые части (23) пропорциональны, а потому уже между $\mathbf{x}_1$ и $\mathbf{x}_2$ имеется линейная зависимость.

Рассмотрим, например, линейное пространство всех многочленов степени $\leqslant 2$ (п. 18), т. е. многочленов вида ax^2+bx+c. Все эти многочлены представляют собой линейные комбинации степеней x^2, x и 1, причем сами эти степени линейно независимы, ни одна из них не является линейной комбинацией остальных. Отсюда сразу следует, что рассматриваемое пространство трехмерно, так как из приведенной леммы вытекает, что любая совокупность, состоящая более чем из трех таких многочленов, линейно зависимая. Указанные степени образуют в этом пространстве базис.

Аналогично получается, что пространство многочленов степени $\leqslant n$ имеет размерность $n+1$. Пространство многочленов всех степеней является бесконечномерным.

Линейное пространство векторов в n-мерном числовом пространстве является, как и следует ожидать, n-мерным. Поясним это на

примере пространства E_4. Для этого введем векторы $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, $\mathbf{e}_4$ со следующими проекциями:

$$\mathbf{e}_1(1;\ 0;\ 0;\ 0),\quad \mathbf{e}_2(0;\ 1;\ 0;\ 0),\quad \mathbf{e}_3(0;\ 0;\ 1;\ 0),\ \mathbf{e}_4(0;\ 0;\ 0;\ 1).$$

Эти векторы линейно независимые, а любой другой вектор в E_4 через них линейно выражается:

$$\mathbf{x}(x_1;\ x_2;\ x_3;\ x_4) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 + x_4\mathbf{e}_4.$$

Отсюда в силу леммы получаем, что рассматриваемое линейное пространство четырехмерно; векторы $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, $\mathbf{e}_4$ образуют в нем базис.

Базис в линейном пространстве можно выбрать бесконечно многими способами. Рассмотрим, например, пространство многочленов степени $\leqslant 2$. Выберем произвольно три значения $x = x_1$, x_2, x_3 и обозначим через $P_k(x)$ $(k = 1,\ 2,\ 3)$ многочлен второй степени, равный 1 при $x = x_k$ и равный нулю при двух остальных значениях x. Нетрудно проверить, что $P_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}$, аналогично выражаются $P_2(x)$ и $P_3(x)$. Эти три многочлена образуют базис в указанном пространстве, а разложение по этому базису — не что иное, как интерполяционная формула Лагранжа (V.23).

В п. 18 мы уже упоминали о понятии линейного подпространства (R_1) линейного пространства (R). Исходя из определения размерности, легко доказать, что если $(R_1) \neq (R)$ и пространство (R) конечномерно, то размерность (R_1) строго меньше размерности (R). При доказательстве этого факта, которое мы предоставляем читателю, полезно опираться на следующее, равносильное применявшемуся ранее, определение линейной зависимости: *векторы* $\mathbf{a}$, $\mathbf{b}$, $\ldots$, $\mathbf{d}$ *называются линейно зависимыми, если между ними имеется линейное соотношение вида* $\alpha\mathbf{a} + \beta\mathbf{b} + \ldots + \delta\mathbf{d} = \mathbf{0}$, *в котором среди чисел* α, β, $\ldots$, δ *по крайней мере одно отлично от нуля.*

С помощью линейных подпространств вводится важное понятие *гиперплоскости* в E_n: гиперплоскость получится, если все векторы, образующие некоторое линейное подпространство в E_n, отложить от некоторой точки E_n и взять концы этих векторов. Гиперплоскости размерности 1 — это прямые, размерности 2 — это плоскости в E_n; имеются также гиперплоскости размерности 3, 4 и т. д. до $n-1$. Формально рассматриваются также «гиперплоскости размерности 0», т. е. точки, и «гиперплоскость размерности n», т. е. все E_n. Можно развить «стереометрию» в E_n.

В заключение укажем на важнейшее понятие *изоморфизма линейных пространств*. Два линейных пространства (R) и (R') называются *изоморфными* (более подробно — *линейно изоморфными*), если векторы из одного пространства можно так сопоставить векторам другого пространства, что линейные действия между соответствующими векторами осуществляются по одинаковым законам. Точнее, это значит, что если векторам $\mathbf{x}$, $\mathbf{y} \in (R)$ соответствуют векторы $\mathbf{x}'$, $\mathbf{y}' \in (R')$, то вектору $\mathbf{x}+\mathbf{y}$ должен соответствовать вектор $\mathbf{x}'+\mathbf{y}'$, а вектору $\lambda\mathbf{x}$ — вектор $\lambda\mathbf{x}'$, т. е.

$$(\mathbf{x}+\mathbf{y})' = \mathbf{x}'+\mathbf{y}',\quad (\lambda\mathbf{x})' = \lambda\mathbf{x}',$$

где штрихом обозначен переход от вектора из (R) к соответствующему вектору из (R'). Конечно, при этом каждому вектору из (R) должен соответствовать один вполне определенный вектор из (R') и обратно. *С точки зрения линейных действий изоморфные пространства неразличимы*, так как любые выводы, которые могут быть сделаны на основе этих действий в одном пространстве, могут быть сделаны и в другом.

Из сказанного вытекает, в частности, что *изоморфные линейные пространства имеют одинаковую размерность*. Обратно, *конечномерные линей-*

ные пространства одинаковой размерности изоморфны. В самом деле, если $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$—базис в (R), а $\mathbf{p}'_1, \mathbf{p}'_2, \ldots, \mathbf{p}'_n$—базис в (R'), то легко проверить, что соответствие $\alpha_1\mathbf{p}_1+\alpha_2\mathbf{p}_2+\ldots+\alpha_n\mathbf{p}_n \leftrightarrow \alpha_1\mathbf{p}'_1+\alpha_2\mathbf{p}'_2+\ldots+\alpha_n\mathbf{p}'_n$ как раз и осуществляет требуемый *изоморфизм* Итак, с точки зрения линейных действий, конечномерные линейные пространства имеют единственный параметр—размерность. В частности, мы получаем, что *каждое конечномерное линейное пространство размерности n изоморфно пространству векторов в E_n.*

20. Понятие евклидова пространства. Линейное пространство (R) называется *евклидовым*, если введено понятие *скалярного произведения* любых двух его векторов, удовлетворяющее следующим естественным аксиомам:

12. Скалярное произведение $\mathbf{x}\cdot\mathbf{y}=(\mathbf{x}, \mathbf{y})$ любых двух векторов $\mathbf{x}, \mathbf{y} \in (R)$ есть вещественное число;

13. $(\mathbf{x}, \mathbf{x})>0$ для любого $\mathbf{x}\neq\mathbf{0}$; 14. $(\mathbf{x}, \mathbf{y})=(\mathbf{y}, \mathbf{x})$;

15. $(\mathbf{x}+\mathbf{y}, \mathbf{z})=(\mathbf{x}, \mathbf{z})+(\mathbf{y}, \mathbf{z})$; 16. $(\lambda\mathbf{x}, \mathbf{y})=\lambda(\mathbf{x}, \mathbf{y})$.

Все эти утверждения для обычных векторов были доказаны в § 3. Таким образом, совокупность обычных векторов образует евклидово пространство. Евклидово пространство образует также совокупность векторов, параллельных некоторой плоскости или параллельных некоторой прямой. Ясно, что всегда линейное подпространство евклидова пространства само является евклидовым пространством.

Подобно п. 18, важнейшим примером евклидова пространства является совокупность всех векторов в n-мерном числовом пространстве E_n, если скалярное произведение любых двух векторов $\mathbf{x}(x_1; x_2; \ldots; x_n)$ и $\mathbf{y}(y_1; y_2; \ldots; y_n)$ определить по формуле

$$(\mathbf{x}, \mathbf{y})=x_1y_1+x_2y_2+\ldots+x_ny_n, \tag{24}$$

аналогичной хорошо известной формуле (12). Легко проверить, что при этом выполняются все аксиомы 12—16.

В евклидовом пространстве вводится понятие длины или, как чаще говорят, *нормы* любого вектора по формуле $|\mathbf{x}|=\sqrt{(\mathbf{x}, \mathbf{x})}$; для нормы применяется также обозначение $\|\mathbf{x}\|$. Из аксиомы 13 видно, что норма любого ненулевого вектора положительна. Если норма вектора равна единице, то этот вектор называется *нормированным* (для обычных векторов употреблялся термин «единичный вектор»). Из аксиом 14 и 16 сразу следует, что

$$|\lambda\mathbf{x}|=\sqrt{(\lambda\mathbf{x}, \lambda\mathbf{x})}=\sqrt{\lambda^2(\mathbf{x}, \mathbf{x})}=|\lambda|\sqrt{(\mathbf{x}, \mathbf{x})}=|\lambda||\mathbf{x}|.$$

Отсюда, в частности, видно, что $|\mathbf{0}|=0$ и что из каждого вектора $\mathbf{x}\neq\mathbf{0}$ можно получить нормированный, разделив $\mathbf{x}$ на $|\mathbf{x}|$; последнее действие называется *нормированием* вектора.

Выведем оценку скалярного произведения $(\mathbf{x}, \mathbf{y})$ двух произвольных векторов $\mathbf{x}, \mathbf{y} \in (R)$. Для этого заметим, что при любом λ будет

$$0 \leqslant |\mathbf{x}+\lambda\mathbf{y}|^2 = (\mathbf{x}+\lambda\mathbf{y}, \mathbf{x}+\lambda\mathbf{y}) = (\mathbf{y}, \mathbf{y})\lambda^2 + 2(\mathbf{x}, \mathbf{y})\lambda + (\mathbf{x}, \mathbf{x}). \quad (25)$$

Если рассмотреть правую часть как квадратный трехчлен относительно λ, то из сохранения его знака вытекает, что дискриминант

$$(\mathbf{x}, \mathbf{y})^2 - (\mathbf{x}, \mathbf{x})(\mathbf{y}, \mathbf{y}) \leqslant 0$$

(почему?). Отсюда получаем

$$|(\mathbf{x}, \mathbf{y})| = \sqrt{(\mathbf{x}, \mathbf{y})^2} \leqslant \sqrt{(\mathbf{x}, \mathbf{x})(\mathbf{y}, \mathbf{y})} = |\mathbf{x}|\,|\mathbf{y}|. \quad (26)$$

Это важное неравенство было выведено в 1821 г. для E_n (в другой терминологии) выдающимся французским математиком О. Коши́ (1789 — 1857), а для некоторых других случаев — в 1859 г. выдающимся русским математиком В. Я. Буняко́вским (1804 — 1889) и в 1884 г. немецким математиком Г. Шварцем (1843 — 1921). Поэтому разные авторы присваивают неравенству (26) разные комбинации из названных фамилий; встречались все восемь мыслимых комбинаций.

Из неравенства (26), в частности, вытекает, что $|\mathbf{x}+\mathbf{y}|^2 = (\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) = (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y}) \leqslant |\mathbf{x}|^2 + 2|\mathbf{x}|\,|\mathbf{y}| + |\mathbf{y}|^2$, откуда

$$|\mathbf{x}+\mathbf{y}| \leqslant |\mathbf{x}| + |\mathbf{y}|.$$

Это неравенство называется *неравенством треугольника* (почему?).

Рассматриваются также *евклидовы пространства над полем комплексных чисел* (п. 17) или, короче, *комплексные евклидовы пространства*. В этом случае скалярное произведение двух векторов будет комплексным числом, а аксиома 14 заменяется на такую: $(x, y) = (y, x)^*$, где звездочкой обозначается комплексно сопряженное число. Отсюда, в частности, вытекает, что $(\mathbf{x}, \lambda\mathbf{y}) = (\lambda\mathbf{y}, \mathbf{x})^* = [\lambda(\mathbf{y}, \mathbf{x})]^* = \lambda^*(\mathbf{x}, \mathbf{y})$. Легко проверить, что прочие утверждения этого пункта остаются в силе. Основным примером является совокупность всех векторов в естественно определенном n-мерном комплексном числовом пространстве Z_n, в котором скалярное произведение, взамен (24), определяется по формуле $(\mathbf{x}, \mathbf{y}) = x_1 y_1^* + x_2 y_2^* + \ldots + x_n y_n^*$.

21. Ортогональность. В евклидовом пространстве (R) естественно вводится понятие *угла* между любыми векторами по формуле

$$\cos(\widehat{\mathbf{x}, \mathbf{y}}) = \frac{(\mathbf{x}, \mathbf{y})}{|\mathbf{x}|\,|\mathbf{y}|}. \quad (27)$$

В силу п. 7 для обычных векторов это определение приводит к обычному углу. Чтобы определение имело смысл, надо заметить, что в силу неравенства (26) правая часть равенства (27) обязательно по абсолютной величине не превосходит единицу. Если векторы $\mathbf{x}, \mathbf{y}$ линейно зависимы, т. е. $\mathbf{y} = \lambda\mathbf{x}$, то из формулы (27) получается, что $(\widehat{\mathbf{x}, \mathbf{y}}) = 0^\circ$, если $\lambda > 0$, и $(\widehat{\mathbf{x}, \mathbf{y}}) = 180^\circ$, если $\lambda < 0$. Из равенства (27) легко получить и обратное, что если $\cos(\widehat{\mathbf{x}, \mathbf{y}}) = \pm 1$,

то векторы $\mathbf{x}$ и $\mathbf{y}$ линейно зависимы; это мы предоставляем читателю.

В наиболее важном случае, когда

$$(\mathbf{x},\ \mathbf{y}) = 0, \tag{28}$$

векторы $\mathbf{x}$ и $\mathbf{y}$ составляют угол $90°$, или, как чаще говорят, являются *ортогональными.* Если хотя бы один из векторов $\mathbf{x}$, $\mathbf{y}$ равен $\mathbf{0}$, то равенство (28) обязательно соблюдается, поэтому *нуль-вектор считается ортогональным любому вектору,* хотя угол, который составляет нуль-вектор с любым заданным вектором, можно считать каким угодно.

Совокупность векторов $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{d}$ называется *ортогональной,* если все эти векторы попарно ортогональны друг другу и ни один из них не равен нуль-вектору. *Такая совокупность всегда линейно независимая,* так как, если равенство $\mathbf{d} = \alpha\mathbf{a} + \beta\mathbf{b} + \ldots$ помножить скалярно на $\mathbf{a}$, мы получаем $0 = \alpha(\mathbf{a},\ \mathbf{a})$, откуда $\alpha = 0$; аналогично получаем, что и все прочие коэффициенты в правой части равны нулю, что невозможно.

В конечномерном евклидовом пространстве наиболее удобно пользоваться *ортогональным базисом,* т. е. таким базисом, который одновременно является ортогональной системой. В самом деле, если $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ — такой базис, так что любой вектор $\mathbf{x}$ можно представить в виде $\mathbf{x} = \alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2 + \ldots + \alpha_n\mathbf{p}_n$, то для нахождения коэффициентов разложения умножим обе части скалярно на любой из векторов $\mathbf{p}_k$. В силу ортогональности получим, что $(\mathbf{x}, \mathbf{p}_k) = \alpha_k(\mathbf{p}_k, \mathbf{p}_k)$, откуда

$$\alpha_k = \frac{(\mathbf{x},\ \mathbf{p}_k)}{(\mathbf{p}_k,\ \mathbf{p}_k)} \quad (k = 1,\ 2, \ldots,\ n). \tag{29}$$

Для неортогонального базиса такой прием привел бы к системе n уравнений первой степени относительно n неизвестных α_k. Формула (29) особенно упрощается, если векторы, составляющие ортогональный базис, являются нормированными (п. 21), тогда в правой части знаменатель отсутствует. Ортогональный базис из нормированных векторов называется *евклидовым базисом* (ср. п. 9).

В любом конечномерном евклидовом пространстве можно выбрать ортогональный базис. В самом деле, пусть, например, пространство (R) четырехмерно и $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$ — какой-то базис в (R). Положим $\mathbf{p}_1 = \mathbf{q}_1$, $\mathbf{p}_2 = \mathbf{q}_2 + \alpha\mathbf{p}_1$ и подберем α так, чтобы $(\mathbf{p}_2, \mathbf{p}_1) = 0$; это даст $(\mathbf{q}_2 + \alpha\mathbf{p}_1, \mathbf{p}_1) = 0$, т. е. $\alpha = -(\mathbf{q}_2, \mathbf{p}_1)(\mathbf{p}_1, \mathbf{p}_1)^{-1}$. Затем положим $\mathbf{p}_3 = \mathbf{q}_3 + \beta_1\mathbf{p}_1 + \beta_2\mathbf{p}_2$ и подберем β_1 и β_2 так, чтобы $(\mathbf{p}_3, \mathbf{p}_1) = 0$ и $(\mathbf{p}_3, \mathbf{p}_2) = 0$; это даст (проверьте!) $\beta_1 = -(\mathbf{q}_3, \mathbf{p}_1)(\mathbf{p}_1, \mathbf{p}_1)^{-1}$, $\beta_2 = -(\mathbf{q}_3, \mathbf{p}_2)(\mathbf{p}_2, \mathbf{p}_2)^{-1}$. Наконец, полагаем $\mathbf{p}_4 = \mathbf{q}_4 + \gamma_1\mathbf{p}_1 + \gamma_2\mathbf{p}_2 + \gamma_3\mathbf{p}_3$ и подбираем коэффициенты γ_k так, чтобы $(\mathbf{p}_4, \mathbf{p}_1) = 0$, $(\mathbf{p}_4, \mathbf{p}_2) = 0$, $(\mathbf{p}_4, \mathbf{p}_3) = 0$; это мы предоставим сделать читателю. Из линейной независимости векторов $\mathbf{q}_k$ легко вывести, что все получающиеся здесь знаменатели отличны от нуля, и поэтому этот *процесс ортогонализации* можно довести до конца; векторы $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4$ образуют в (R) ортогональный базис.

Отсюда вытекает важное следствие об изоморфизме евклидовых пространств. При том два е[illegible] до ых прос ран а называются *евклидово*

изоморфными, если между ними можно установить линейный изоморфизм (п. 19), сохраняющий скалярное произведение, т. е. $(\mathbf{x}, \mathbf{y})_R = (\mathbf{x}', \mathbf{y}')_{R'}$ (буква внизу указывает, в каком пространстве берется скалярное произведение). С точки зрения теории евклидовых пространств два евклидово изоморфных пространства неразличимы. Ясно, что два евклидово изоморфных пространства имеют одинаковую размерность. Однако аналогично п. 19 легко установить и обратное, что два евклидовых пространства, имеющих одинаковую конечную размерность, евклидово изоморфны. Для этого достаточно в каждом из этих пространств выбрать ортогональный базис, состоящий из нормированных векторов (если ортогональные векторы пронормировать, то ортогональность не нарушится). Пусть это базисы $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ и $\mathbf{p}_1', \mathbf{p}_2', \ldots, \mathbf{p}_n'$; тогда требуемый изоморфизм осуществляется по формуле $\alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2 + \ldots \ldots + \alpha_n\mathbf{p}_n \leftrightarrow \alpha_1\mathbf{p}_1' + \alpha_2\mathbf{p}_2' + \ldots + \alpha_n\mathbf{p}_n'$; в самом деле, если $\mathbf{x} = \alpha_1\mathbf{p}_1 + \ldots + \alpha_n\mathbf{p}_n$, $\mathbf{y} = \beta_1\mathbf{p}_1 + \ldots + \beta_n\mathbf{p}_n$, то $(\mathbf{x}, \mathbf{y})_R = \alpha_1\beta_1 + \ldots + \alpha_n\beta_n = (\mathbf{x}', \mathbf{y}')_{R'}$ (почему?). В частности, мы получаем, что *каждое конечномерное евклидово пространство размерности n евклидово изоморфно пространству векторов в E_n*. Аналогичное утверждение верно и для комплексных пространств.

§ 7. Векторные функции скалярного аргумента и кривизна

22. Переменные векторные величины. Вернемся вновь к обычным векторам. Переменные векторные величины тесно связаны с переменными скалярными величинами, так как если, например, ввести в пространстве декартовы координаты и записать

$$\mathbf{u} = u_x\mathbf{i} + u_y\mathbf{j} + u_z\mathbf{k}, \tag{30}$$

то видно, что всякое изменение векторной величины $\mathbf{u}$ сводится к изменению трех скалярных величин u_x, u_y, u_z. Сделаем только несколько простых замечаний.

Векторная величина $\mathbf{u}$ является *бесконечно малой*, $\mathbf{u} \to \mathbf{0}$, если ее модуль u бесконечно мал, $u \to 0$; при этом направление вектора $\mathbf{u}$ может меняться произвольно и не обязано иметь предел.

Векторное предельное соотношение $\mathbf{u} \to \mathbf{a}$ ($\mathbf{a}$ — постоянный вектор) равносильно трем скалярным: $u_x \to a_x$; $u_y \to a_y$; $u_z \to a_z$. Если при этом $\mathbf{a} \neq \mathbf{0}$, то $\mathbf{u}$ стремится к $\mathbf{a}$ как по модулю, так и по направлению. Теоремы о пределе суммы, произведения и т. п. (см. п. III.5) сохраняются вместе с доказательствами. Однако свойства пределов, связанные с неравенствами, на векторные величины, конечно, не распространяются, так как векторные величины знаками неравенства соединять нельзя.

23. Векторная функция скалярного аргумента. *Векторная функция скалярного аргумента* $\mathbf{u} = \mathbf{f}(t)$ *получается, если каждому значению скалярной величины* t *по определенному закону ставится в соответствие некоторое значение векторной величины* $\mathbf{u}$. Подобно началу п. 22, задание одной векторной функции равносильно заданию трех скалярных (обычных) функций, так как $u_x = f_x(t)$, $u_y = f_y(t)$ и $u_z = f_z(t)$.

Понятие непрерывности векторной функции скалярного аргумента вводится так же, как и для скалярной функции. Затем

$$\mathbf{u}' = \mathbf{f}'(t) = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{u}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\mathbf{f}(t+\Delta t) - \mathbf{f}(t)}{\Delta t}.$$

Основные свойства производной (п. IV.4) переносятся вместе с их доказательствами; подчеркнем только, что в правой части формулы для производной векторного произведения $(\mathbf{u} \times \mathbf{v})' = \mathbf{u}' \times \mathbf{v} + \mathbf{u} \times \mathbf{v}'$ надо следить за порядком множителей. Переносится также ряд Тейлора (IV.52).

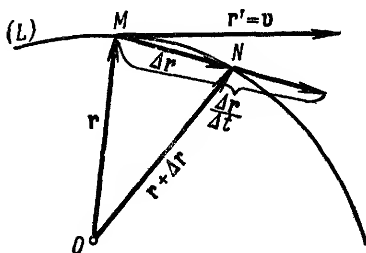


Рис. 172.

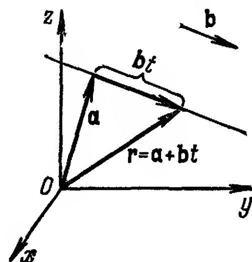


Рис. 173.

Иногда применяется следующее свойство: если $|\mathbf{u}(t)| = \text{const}$, то $\mathbf{u}' \perp \mathbf{u}$. Действительно, условие можем записать так: $\mathbf{u} \cdot \mathbf{u} = u^2 = \text{const}$, и, продифференцировав, получим $\mathbf{u}' \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{u}' = 0$, $2\mathbf{u}' \cdot \mathbf{u} = 0$, $\mathbf{u}' \perp \mathbf{u}$.

Чтобы придать геометрический смысл векторной функции скалярного аргумента, будем откладывать вектор $\mathbf{u}$ от некоторой фиксированной точки O пространства; тогда естественно $\mathbf{u}$ называть радиусом-вектором (см. п. 9) и обозначать буквой $\mathbf{r}$, т. е.

$$\mathbf{r} = \mathbf{f}(t). \tag{31}$$

При изменении t конец вектор $\mathbf{r}$ опишет в пространстве некоторую линию (L) (рис. 172), так что (31) можно считать *векторно-параметрическим уравнением линии* (L). Впрочем, если ввести декартовы оси, то от (31) легко перейти к скалярно-параметрическим уравнениям

$$x = \varphi(t), \quad y = \psi(t), \quad z = \chi(t) \tag{32}$$

той же линии; здесь правые части получаются в результате проектирования функции $\mathbf{f}(t)$ на оси координат. (Ср. параметрические уравнения (II.10) линии на плоскости.) Обратный переход от (32) к (31) совершается по формуле $\mathbf{r} = \varphi(t)\,\mathbf{i} + \psi(t)\,\mathbf{j} + \chi(t)\,\mathbf{k}$. Переменную t удобно истолковывать как время. Тогда линия (L) будет представлять собой траекторию движущейся точки. Эта линия называется также *годóграфом* векторной величины $\mathbf{u}$.

Например, как показано на рис. 173, уравнение $\mathbf{r} = \mathbf{a} + \mathbf{b}t$, где $\mathbf{a}$ и $\mathbf{b}$ — постоянные векторы, определяет прямую линию и равномерное

движение по ней со скоростью $\mathbf{b}$. Проектируя это уравнение на оси координат, получим *параметрические уравнения прямой линии*:

$$x = a_x + b_x t, \quad y = a_y + b_y t, \quad z = a_z + b_z t. \qquad (33)$$

В качестве другого примера рассмотрим уравнения *винтовой линии*. Эта линия получается в результате наложения равномерного перемещения параллельно некоторой оси и равномерного вращения вокруг этой же оси (рис. 174). Приняв в качестве этой оси ось z и обозначив поступательную скорость через v, а вращательную через ω, получим

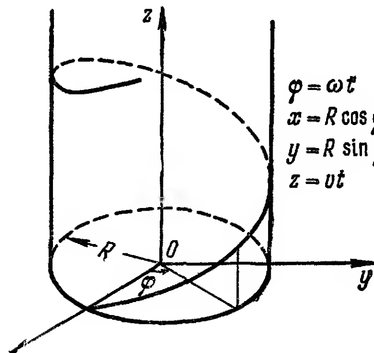


Рис. 174.

$$x = R\cos\omega t; \quad y = R\sin\omega t; \quad z = vt$$

или в векторном виде $\mathbf{r} = R(\cos\omega t\ \mathbf{i} + \sin\omega t\ \mathbf{j}) + vt\ \mathbf{k}$.

Если переменная t получит приращение Δt, то точка M на линии (L) перейдет в положение N (см. рис. 172); таким образом, $\Delta\mathbf{r} = \overrightarrow{MN}$. Отношение $\Delta\mathbf{r}/\Delta t$ (*вектор средней скорости*), поскольку Δt — скаляр, также идет по прямой MN. Когда $\Delta t \to 0$, то $\Delta\mathbf{r}/\Delta t \to \mathbf{r}' = \mathbf{v}$ (*вектор мгновенной скорости*); но секущая MN, будучи продолжена, при этом (т. е. когда $N \to M$), поворачиваясь около точки M, стремится перейти в касательную к траектории в точке M. Таким образом, мы видим, что *вектор мгновенной скорости* $\mathbf{v} = d\mathbf{r}/dt$ *направлен по касательной к траектории*, что, впрочем, хорошо известно из механики. Векторно-параметрическое уравнение касательной, проведенной при некотором значении $t = t_0$, имеет вид (см. рис. 173)

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0(t - t_0),$$

где $\mathbf{r}_0 = \mathbf{f}(t_0)$; $\mathbf{v}_0 = \mathbf{f}'(t_0)$. Это уравнение можно переписать так: $\Delta\mathbf{r} = d\mathbf{f}$ вместо $\Delta\mathbf{r} = \Delta\mathbf{f}$ для линии (L). Значит, замена $\Delta\mathbf{f}$ на $d\mathbf{f}$ означает переход от движения по линии (L) к равномерному движению по касательной со скоростью, равной мгновенной скорости в рассматриваемый момент, как будто на движущуюся точку вдруг перестали действовать все силы.

Согласно п. III.9 (пример 1) при $\Delta s \to 0$

$$\frac{|\Delta s|}{|\Delta\mathbf{r}|} \to 1; \quad \left|\frac{\Delta\mathbf{r}}{\Delta s}\right| \to 1; \quad \left|\frac{d\mathbf{r}}{ds}\right| = \lim\left|\frac{\Delta\mathbf{r}}{\Delta s}\right| = 1; \quad |d\mathbf{r}| = |ds|; \qquad (34)$$

Δs и ds берутся по абсолютной величине, так как эти величины могут быть как положительными, так и отрицательными. Таким образом, если за параметр вдоль (L) взять длину s дуги, отсчитываемую от некоторой точки M_0 этой линии, т. е. положить $\mathbf{r} = \mathbf{r}(s)$,

то производная $d\mathbf{r}/ds$ даст единичный вектор, направленный по касательной к (L), т. е. орт касательной (см. п. 7). Этот вектор принято обозначать буквой $\boldsymbol{\tau}$, т. е. $\frac{d\mathbf{r}}{ds}=\boldsymbol{\tau}$,

$$\frac{d\mathbf{r}}{dt}=\mathbf{v}=\frac{d\mathbf{r}}{ds}\frac{ds}{dt}=v\boldsymbol{\tau};\qquad \boldsymbol{\tau}=\frac{d\mathbf{r}/dt}{ds/dt}=\frac{\dot{\mathbf{r}}}{\dot{s}}. \tag{35}$$

Отметим в заключение, что из формулы (34) вытекает выражение для дифференциала дуги в декартовых координатах:

$$ds=\pm|d\mathbf{r}|=\pm|d(x\mathbf{i}+y\mathbf{j}+z\mathbf{k})|=$$
$$=\pm|dx\mathbf{i}+dy\mathbf{j}+dz\mathbf{k}|=\pm\sqrt{dx^2+dy^2+dz^2}.$$

24. Понятия, связанные со второй производной. Так как $|\boldsymbol{\tau}(s)|=1=\text{const}$, то (см. начало п. 23) $d\boldsymbol{\tau}/ds\perp\boldsymbol{\tau}$. Таким образом, прямая pp (рис. 175), проведенная через некоторую текущую точку M траектории (L) параллельно $d\boldsymbol{\tau}/ds$, будет служить *нормалью* к (L), т. е. перпендикуляром к касательной ll в точке M. Чтобы отличить ее от остальных нормалей (а в каждой точке линии в пространстве можно провести к ней бесконечное число нормалей, которые заполняют *нормальную плоскость*), ее называют *главной нормалью* в точке M к линии (L). Модуль вектора $d\boldsymbol{\tau}/ds$ называется *кривизной линии* (L) в точке M и обозначается буквой k, т. е. $\left|\frac{d\boldsymbol{\tau}}{ds}\right|=k$; $\frac{d\boldsymbol{\tau}}{ds}=k\mathbf{n}$, где $\mathbf{n}$ — орт главной нормали.

Рис. 175.

Рис. 176.

Из рис. 176 виден геометрический смысл кривизны:

$$k=\left|\frac{d\boldsymbol{\tau}}{ds}\right|=\lim\left|\frac{\Delta\boldsymbol{\tau}}{\Delta s}\right|=\lim\frac{BC}{\Delta s}=$$
$$=\lim\frac{2\sin\frac{\alpha}{2}}{\Delta s}=\lim\frac{\alpha}{\Delta s};$$

в последнем переходе мы воспользовались свойством 4 п. III.8. Таким образом, *кривизна есть скорость поворота касательной в расчете на единицу длины пройденного пути*. Заодно видно, что

вектор $\boldsymbol{\tau}'_s$, а с ним и $\mathbf{n}$ направлены в сторону вогнутости кривой.

Дифференцируя первую формулу (35) по t, получим

$$\frac{d^2\mathbf{r}}{dt^2}=\frac{dv}{dt}\boldsymbol{\tau}+v\frac{d\boldsymbol{\tau}}{dt}=\frac{dv}{dt}\boldsymbol{\tau}+v\frac{d\boldsymbol{\tau}}{ds}\frac{ds}{dt}=\frac{dv}{dt}\boldsymbol{\tau}+v^2k\mathbf{n}.$$

Эта формула широко применяется в механике, так как если t является временем, формула дает разложение вектора ускорения на *тангенциáльную*, т. е. направленную по касательной, и *нормальную составляющие*; последняя, как видим, направлена именно по главной нормали.

Таким образом, вектор $d^2\mathbf{r}/dt^2$, будучи отложен от точки M, обязательно лежит в плоскости, проходящей через касательную и главную нормаль, проведенные в этой точке; эта плоскость называется *соприкасающейся плоскостью* к линии (L) в точке M. Из формулы Тейлора (IV.50), примененной к $\mathbf{f}(t_0+\Delta t)$, видим, что с точностью до величины третьего порядка относительно Δt линию (L) можно считать лежащей в своей соприкасающейся плоскости. (Подобным образом, из формулы (IV.49) следует, что с точностью до величин второго порядка малости линию (L) можно считать совпадающей со своей касательной.) Отсюда можно было бы сделать вывод, что *соприкасающаяся плоскость к линии (L) в точке M — это плоскость, проходящая через три точки линии (L), расположенные в бесконечной близости от M* (подобно тому как *касательная — это прямая, проведенная через две бесконечно близкие точки линии*).

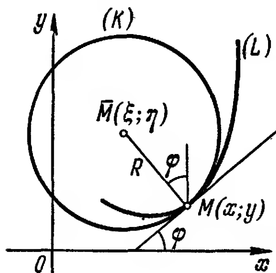


Рис. 177.

$\operatorname{tg}\varphi=\frac{dy}{dx}=\frac{\dot y}{\dot x}$ (см. п. IV 9).

25. Соприкасающаяся окружность. Пусть линия (L) *на плоскости* задана в параметрическом виде (см. п. II.6): $x=x(t)$; $y=y(t)$. Тогда из рис. 177 следует, что ее кривизна k в любой текущей точке равна $k=\left|\frac{d\varphi}{ds}\right|=\left|\frac{d\operatorname{arctg}\frac{\dot y}{\dot x}}{ds}\right|$, где точкой сверху обозначается производная по параметру. Пользуясь выражением для ds из конца п. 23 и проводя преобразования, получим

$$k=\left|\frac{1}{1+\left(\frac{\dot y}{\dot x}\right)^2}\frac{\ddot y\dot x-\dot y\ddot x}{\dot x^2}\,dt\right|:\sqrt{dx^2+dy^2}=$$

$$=\left|\frac{1}{\dot x^2+\dot y^2}(\ddot y\dot x-\dot y\ddot x)\,dt\right|:\sqrt{\dot x^2+\dot y^2}\,dt=\frac{|\ddot y\dot x-\dot y\ddot x|}{(\dot x^2+\dot y^2)^{3/2}}. \qquad (36)$$

В частности, если уравнение линии дано в виде $y=f(x)$, т. е. за параметр взят сам x, то $\dot y=y'$, $\ddot y=y''$, $\dot x=1$, $\ddot x=0$ и

$$k=\frac{|y''|}{(1+y'^2)^{3/2}}. \qquad (37)$$

Из рис. 178 видно, что *для окружности*

$$k=\left|\frac{d\varphi}{ds}\right|=\lim\left|\frac{\Delta\varphi}{\Delta s}\right|=\lim\frac{\Delta\varphi}{R\Delta\varphi}=\frac{1}{R}, \tag{38}$$

т. е. *кривизна постоянна и обратна радиусу*. Из плоских кривых только у окружности и у прямой кривизна постоянная; у прямой она равна нулю.

Возьмем произвольную точку M на линии (L); окружность (K), проходящая через M, имеющая там такую же касательную и такую же кривизну, что и (L), и выгнутая в ту же сторону, что (L), называется *соприкасающейся окружностью* («*кругом кривизны*») для линии (L) в точке M (см. рис. 177). Ее радиус называется *радиусом кривизны*, а центр называется *центром кривизны* линии (L) в точке M. Согласно формулам (36) — (38)

$$R=\frac{1}{k_{\text{сопр. окружн}}}=\frac{1}{k_{\text{линии }(L)}}=\left|\frac{ds}{d\varphi}\right|=\frac{(\dot{x}^2+\dot{y}^2)^{3/2}}{|\ddot{y}\dot{x}-\dot{y}\ddot{x}|}=\frac{(1+y'^2)^{3/2}}{|y''|}. \tag{39}$$

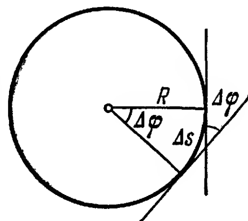


Рис. 178.
$\Delta s=R\Delta\varphi$.

Линия (L) вблизи M весьма тесно прилегает к своей соприкасающейся окружности. С помощью формулы Тейлора (IV.50) можно было бы вывести, что с точностью до величин третьего порядка относительно Δt линию (L) можно считать совпадающей со своей соприкасающейся окружностью. Огсюда в свою очередь можно вывести, что соприкасающаяся окружность — это окружность, проходящая через три бесконечно близкие точки линии (L). Последнее определение пригодно и для пространственных линий; из него, в частности, сразу следует, что соприкасающаяся окружность лежит в соприкасающейся плоскости.

Точки кривой, в которых кривизна имеет экстремальное значение (но не точки перегиба!), называются *вершинами* этой кривой. Например, для эллипса с параметрическими уравнениями $x=a\cos t$, $y=b\sin t$

$$k=\frac{|(-b\sin t)(-a\sin t)-(b\cos t)(-a\cos t)|}{[(-a\sin t)^2+(b\cos t)^2]^{3/2}}=\frac{ab}{(a^2\sin^2 t+b^2\cos^2 t)^{3/2}}.$$

Однако знаменатель имеет экстремальные значения при $t=0$, $\frac{\pi}{2}$, π, $\frac{3}{2}\pi$ и т. д. (проверьте!). Поэтому вершинами эллипса в новом смысле являются те самые вершины, которые были определены в п. II.10. Радиус кривизны в вершинах равен b^2/a в точках пересечения эллипса с осью x и a^2/b в точках пересечения с осью y. Этот результат применяется для приближенного построения эллипса: циркулем наводятся круги кривизны в вершинах (рис. 179), после чего переход от одного круга к другому осуществляется с помощью лекала. На рис. 179 для простоты показана только четверть эллипса; показано также построение радиусов кривизны в вершинах.

Для нахождения координат ξ, η центра кривизны линии (L) в точке M положим, что эта линия в M выпукла вниз, т. е. $y''>0$ (см. п. IV.20), как

на рис. 177. Тогда

$$\left.\begin{aligned}\xi&=x-R\sin\varphi=x-R\frac{\operatorname{tg}\varphi}{\sqrt{1+\operatorname{tg}^2\varphi}}=x-\frac{(1+y'^2)^{3/2}}{y''}\,\frac{y'}{\sqrt{1+y'^2}}=\\&\qquad\qquad=x-\frac{y'(1+y'^2)}{y''};\\\eta&=y+R\cos\varphi=y+R\frac{1}{\sqrt{1+\operatorname{tg}^2\varphi}}=\\&\qquad\qquad=y+\frac{(1+y'^2)^{3/2}}{y''}\,\frac{1}{\sqrt{1+y'^2}}=y+\frac{1+y'^2}{y''}.\end{aligned}\right\}\tag{40}$$

Эти же формулы получаются, если $y''<0$. Переход (проделайте его!) к производным по параметру дает

$$\xi=x-\frac{\dot y(\dot x^2+\dot y^2)}{\ddot y\dot x-\dot y\ddot x};\quad \eta=y+\frac{\dot x(\dot x^2+\dot y^2)}{\ddot y\dot x-\dot y\ddot x}.\tag{41}$$

26. Эволюта и эвольвента. *Геометрическое место* $(\bar L)$ *центров кривизны линии* (L) *называется эволю́той этой линии*; в свою очередь *линия называется эвольве́нтой по отношению к своей эволюте*, т. е. если $(\bar L)$ является эволютой линии (L), то (L) является эвольвентой линии $(\bar L)$ и наоборот. Можно доказать, что эволюта связана с эвольвентой следующим образом:

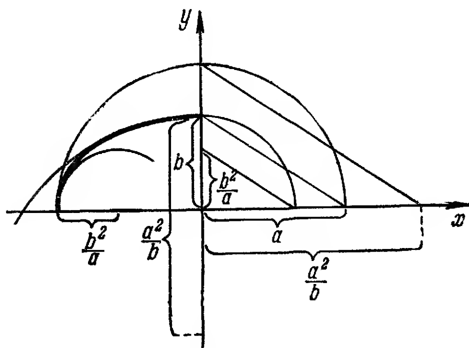


Рис. 179.

1) если M — любая точка эвольвенты, а $\bar M$ — соответствующая точка эволюты, т. е. $\bar M$ является центром кривизны линии (L) в точке M, то *прямая* $M\bar M$ *служит не только нормалью к эвольвенте, но и касательной к эволюте* (рис. 180);

2) если точка M движется по эвольвенте, то *изменение радиуса кривизны равно длине дуги эволюты между соответствующими центрами кривизны* (рис. 181).

Из этих свойств следует, что *эвольвента* (от латинского evolvo, что значит «развертываю», «разворачиваю») *получается в результате сматывания натянутой нити с контура, имеющего форму эволюты.* Особо важное значение имеет эвольвента (развертка) окружности, так как профили подавляющего большинства зубьев у зубчатых колес очерчены с боков дугами этой эвольвенты.

Если уравнение линии (L) дано в параметрическом виде $x=x(t)$, $y=y(t)$, то параметрические уравнения ее эволюты $\xi=\xi(t)$, $\eta=\eta(t)$ (где ξ и η откладываются по тем же осям, что x и y) получается из формул (41), если в них подставить выражения x и y через t.

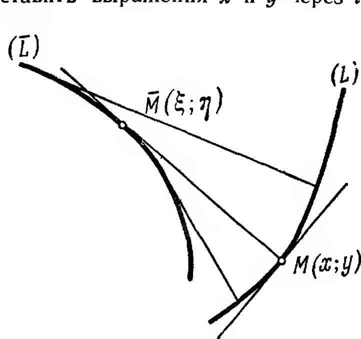


Рис. 180.
$(\bar{L})$ — эволюта; (L) — эвольвента.

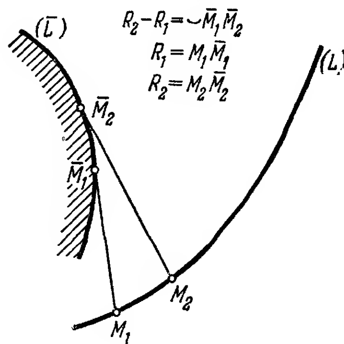


Рис. 181.

Для доказательства указанных двух свойств продифференцируем начала равенств (40), справедливые, если $d\varphi/ds>0$:

$$d\xi=dx-dR\sin\varphi-R\cos\varphi\,d\varphi;\quad d\eta=dy+dR\cos\varphi-R\sin\varphi\,d\varphi. \tag{42}$$

Однако согласно формуле (39)

$$dx=ds\cos\varphi=\frac{ds}{d\varphi}\cos\varphi\,d\varphi=R\cos\varphi\,d\varphi,\quad \text{аналогично } dy=R\sin\varphi\,d\varphi. \tag{43}$$

Подставляя (43) в (42), получим

$$d\xi=-dR\sin\varphi;\ d\eta=dR\cos\varphi; \tag{44}$$

тот же окончательный результат получается, если $d\varphi/ds<0$.

Из формул (44) прежде всего следует, что $\frac{d\eta}{d\xi}=-\operatorname{ctg}\varphi=-\frac{1}{dy/dx}$. Это значит (см. задачу 5 п. II.9), что касательная к эвольвенте в любой ее точке M и касательная к эволюте в соответствующей точке $\bar{M}$ взаимно перпендикулярны (см. рис. 180), откуда и следует первое из двух указанных выше свойств.

Из формул (44) получается также $d\xi^2+d\eta^2=dR^2$. Отсюда (см. конец п. 23) следует, что $dR=\pm ds$, где s — длина дуги, отсчитываемая вдоль эволюты. Значит, $d(R\mp s)=0$; $R\mp s=\text{const}$; $R=C\pm s$; $\Delta R=\pm\Delta s$; $|\Delta R|=|\Delta s|$. Этим доказывается и второе из указанных свойств эволюты и эвольвенты.

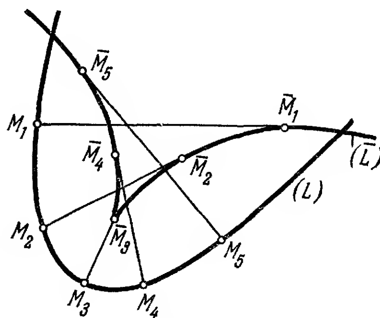


Рис. 182

Можно показать, что вершинам линии (см. п. 25) отвечают точки возврата ее эволюты (рис. 182). Например, у эволюты эллипса четыре точки возврата. Если линия в какой-либо точке имеет нулевую кривизну (так, в частности, обычно бывает в точках перегиба), то соответствующая точка эволюты уходит в бесконечность.

В качестве примера найдем эволюту циклоиды [см. формулы (II.12), в которых вместо ψ надо писать t]. Здесь

$$x=R(t-\sin t);\quad \dot{x}=R(1-\cos t);\quad \ddot{x}=R\sin t;$$
$$y=R(1-\cos t);\quad \dot{y}=R\sin t;\quad \ddot{y}=R\cos t;$$
$$\dot{x}^2+\dot{y}^2=R^2(1-\cos t)^2+R^2\sin^2 t=2R^2(1-\cos t);$$
$$\ddot{y}\dot{x}-\dot{y}\ddot{x}=R\cos t\cdot R(1-\cos t)-R\sin t\cdot R\sin t=-R^2(1-\cos t).$$

По формулам (41) получим

$$\xi=R(t-\sin t)-\frac{R\sin t\cdot 2R^2(1-\cos t)}{-R^2(1-\cos t)}=R(t+\sin t);$$

аналогично $\eta=-R(1-\cos t)$. Если обозначить $t=\pi+\tau$, то получим

$$\xi=R[\pi+\tau+\sin(\pi+\tau)]=R(\tau-\sin\tau)+\pi R;$$
$$\eta=-R[1-\cos(\pi+\tau)]=R(1-\cos\tau)-2R.$$

Так как πR и $2R$ — константы, то, сравнивая с формулами (II. 12), видим, что эволюта циклоиды — это циклоида тех же размеров, но сдвинутая относительно первой на πR вдоль оси x и на $-2R$ вдоль оси y. Этот результат применяется в механике.

КОМПЛЕКСНЫЕ ЧИСЛА И ФУНКЦИИ

Комплексные числа широко применяются в современной математике и ее приложениях. Оказывается, что многие соотношения между вещественными (действительными) величинами удобно получать, используя по пути комплексные числа и комплексные функции.

§ 1. Комплексные числа

1. Комплексная плоскость. Определение комплексных чисел известно из средней школы: *комплексным числом* называется агрегат вида

$$z = x + iy, \tag{1}$$

где x и y — вещественные числа, а i — *мнимая единица*, удовлетворяющая равенству $i^2 = -1$. При этом x называется *вещественной частью*, а y — *мнимой частью* числа z, что записывается так: $x = \operatorname{Re} z$, $y = \operatorname{Im} z$. (Иногда мнимой частью числа z называют все произведение iy, что более естественно, но менее удобно.) *Два комплексных числа равны тогда и только тогда, если равны их вещественные, а также мнимые части*: если $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$, то равенства

$$z_1 = z_2 \quad \text{и} \quad \left.\begin{array}{l} x_1 = x_2, \\ y_1 = y_2 \end{array}\right\} \tag{2}$$

равносильны. Таким образом, *одно «комплексное равенство» равносильно двум вещественным. Знаками неравенства комплексные числа соединять нельзя*, т. е. неравенств вида $z_1 > z_2$ не существует.

Комплексные числа изображаются на плоскости. Для этого выбирают систему декартовых координат x, y, после чего любое число вида (1) изображается точкой $M(x; y)$. Такая плоскость условно называется *комплексной плоскостью*, хотя, конечно, все точки на ней имеют вещественные координаты. При этом часто для сокращения вместо «точка, отвечающая числу $x + iy$» говорят просто «точка $x + iy$» (рис. 183).

Вещественные числа являются частным случаем комплексных, если в формуле (1) положить $y = 0$; они изображаются точками на *вещественной оси*, т. е. оси x. Комплексные невещественные числа называются *мнимыми*; таким образом, *всякое комплексное число является либо вещественным, либо мнимым.* Если у комплексного

числа вещественная часть отсутствует (равна нулю), то оно называется *чисто мнимым*; такие числа изображаются на *мнимой оси*, т. е. оси y.

На комплексной плоскости часто рассматриваются также полярные координаты (см. рис. 183 и п. II.3); они называются *модулем и аргументом комплексного числа* z и обозначаются $\rho = |z|$, $\varphi = \arg z$. Как известно, $\rho = \sqrt{x^2 + y^2}$, $x = \rho \cos \varphi$, $y = \rho \sin \varphi$, откуда в силу (1)

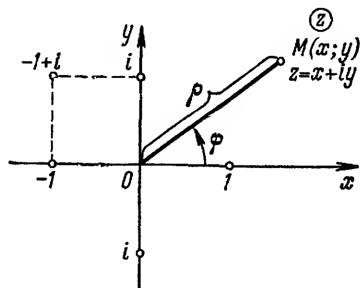


Рис. 183.

$$z = \rho (\cos \varphi + i \sin \varphi). \quad (3)$$

Таким образом, любое комплексное число можно записать в *тригонометрической форме* (3). Модуль любого комплексного числа имеет вполне определенное значение, тогда как аргумент определен с точностью до целого числа полных оборотов; например, $|i| = 1$, $\operatorname{Arg} i = \frac{\pi}{2} + 2k\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$. Здесь под Arg понимается *общее значение аргумента*, составленное с учетом возможности любого числа полных оборотов; в отличие от этого *главное значение* arg берется в интервале $-180° < \arg \leqslant 180°$. Числу $z = 0$ может быть приписан любой аргумент.

2. Алгебраические действия над комплексными числами. При сложении комплексных чисел складываются их вещественные части, а также их мнимые части. Если сравнить это правило с правилом сложения векторов [см. формулу (VII.10)], то видно, что *комплексные числа складываются и вычитаются, как векторы* (рис. 184). Отсюда, в частности, следует, что

$$|z_1 + z_2| \leqslant |z_1| + |z_2|.$$

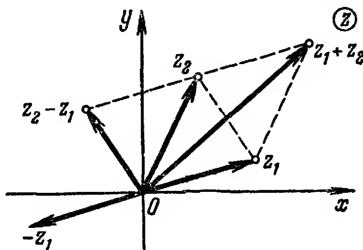


Рис. 184.

Аналогично проверяется, что *комплексные числа множатся на вещественные, как векторы на скаляры*. На основании этих свойств комплексные числа часто истолковываются как *векторы в комплексной плоскости*. Переход от изображения комплексных чисел векторами к изображению точками на плоскости очевиден: надо вектор отложить от начала координат, тогда его конец окажется в соответствующей точке.

Умножение комплексных чисел совершенно отлично от умножения векторов. Если воспользоваться тригонометрической формой

$$z_1 = \rho_1 (\cos \varphi_1 + i \sin \varphi_1), \quad z_2 = \rho_2 (\cos \varphi_2 + i \sin \varphi_2),$$

то произведение $z=z_1 \cdot z_2$ можно записать в виде

$$z=\rho(\cos\varphi+i\sin\varphi)=\rho_1(\cos\varphi_1+i\sin\varphi_1)\,\rho_2(\cos\varphi_2+i\sin\varphi_2)=$$
$$=\rho_1\rho_2[\cos\varphi_1\cos\varphi_2+i\cos\varphi_1\sin\varphi_2+i\sin\varphi_1\cos\varphi_2-\sin\varphi_1\sin\varphi_2]=$$
$$=\rho_1\rho_2[\cos(\varphi_1+\varphi_2)+i\sin(\varphi_1+\varphi_2)].$$

Таким образом,

$$\rho=\rho_1\rho_2, \quad \varphi=\varphi_1+\varphi_2,$$

т. е.

$$|z_1z_2|=|z_1|\cdot|z_2|, \quad \operatorname{Arg}(z_1\cdot z_2)=\operatorname{Arg} z_1+\operatorname{Arg} z_2;$$

при умножении комплексных чисел их модули перемножаются, а аргументы складываются. Отсюда при обратном действии, делении,

$$\left|\frac{z_1}{z_2}\right|=\frac{|z_1|}{|z_2|}, \quad \operatorname{Arg}\frac{z_1}{z_2}=\operatorname{Arg} z_1-\operatorname{Arg} z_2.$$

Особенно интересно умножение комплексного числа на i: так как $|i|=1$, $\arg i=\frac{\pi}{2}$, то $|iz|=|z|$, $\operatorname{Arg} iz=\operatorname{Arg} z+\frac{\pi}{2}$; значит, вектор iz получается из вектора z поворотом в положительном направлении на 90°.

Правило умножения комплексных чисел автоматически распространяется на любое число сомножителей. Если, в частности, взять все сомножители равными, то получаем

$$[\rho(\cos\varphi+i\sin\varphi)]^n=\rho^n(\cos n\varphi+i\sin n\varphi) \qquad (n=2,\ 3,\ \ldots);$$

если же положить $\rho=1$, получим формулу

$$(\cos\varphi+i\sin\varphi)^n=\cos n\varphi+i\sin n\varphi,$$

названную *формулой Муавра* по имени английского математика А. Муа́вра (1667—1754), нашедшего ее в 1707 г.

Формулу Муавра можно применить для выражения тригонометрических функций кратных дуг; например, положив $n=3$, получим

$$(\cos\varphi+i\sin\varphi)^3=\cos^3\varphi+i3\cos^2\varphi\sin\varphi-$$
$$-3\cos\varphi\sin^2\varphi-i\sin^3\varphi=\cos 3\varphi+i\sin 3\varphi,$$

т. е. согласно формуле (2)

$$\cos 3\varphi=\cos^3\varphi-3\cos\varphi\sin^2\varphi, \quad \sin 3\varphi=3\cos^2\varphi\sin\varphi-\sin^3\varphi.$$

Конечно, при этом надо иметь в виду таблицу степеней числа i: $i^1=i$, $i^2=-1$, $i^3=-i$, $i^4=1$, $i^5=i$, $i^6=-1$ и т. д. Отметим, кстати, что $\frac{1}{i}=-i$.

Переходим к извлечению корня из комплексных чисел. Если $z=\rho(\cos\varphi+i\sin\varphi)$ дано, $\sqrt[n]{z}=w=r(\cos\psi+i\sin\psi)$ ищется, то по определению корня $z=w^n=r^n(\cos n\psi+i\sin n\psi)$. Сравнивая с предыдущим выражением для z, получим (см. конец п. 1):

$$r^n=\rho, \quad n\psi=\varphi+2k\pi \qquad (k\text{ — любое целое}).$$

Так как r и ρ — положительные числа, то $r=(\sqrt[n]{\rho})_{об}$, $\psi=\frac{\varphi+2k\pi}{n}$, где под

$(\sqrt[n]{\rho})_{об}$ понимается «обыкновенный» (арифметический, положительный) корень из положительного числа. Таким образом,

$$w=(\sqrt[n]{\rho})_{об}\left(\cos\frac{\varphi+2k\pi}{n}+i\sin\frac{\varphi+2k\pi}{n}\right).$$

Придавая k значения 0, 1, 2, ..., мы будем получать возможные значения для корня w_1, w_2, w_3, ... Однако при $k=n$ получится

$$w_{n+1}=(\sqrt[n]{\rho})_{об}\left(\cos\frac{\varphi+2n\pi}{n}+i\sin\frac{\varphi+2n\pi}{n}\right)=$$
$$=(\sqrt[n]{\rho})_{об}\left[\cos\left(\frac{\varphi}{n}+2\pi\right)+i\sin\left(\frac{\varphi}{n}+2\pi\right)\right]=$$
$$=(\sqrt[n]{\rho})_{об}\left(\cos\frac{\varphi}{n}+i\sin\frac{\varphi}{n}\right)=w_1.$$

Аналогично, $w_{n+2}=w_2$ и т. д.; при отрицательных k также не будет ничего нового: при $k=-1$ получится то же, что при $k=n-1$, и т. д. Окончательно,

$$\left[\sqrt[n]{\rho(\cos\varphi+i\sin\varphi)}\right]_{k+1}=$$
$$=(\sqrt[n]{\rho})_{об}\left(\cos\frac{\varphi+2k\pi}{n}+i\sin\frac{\varphi+2k\pi}{n}\right)\qquad(k=0,\ 1,\ \ldots,\ n-1).\quad(4)$$

В частности, мы видим, что *корень n-й степени из комплексного числа имеет n различных значений*; единственным исключением является число $z=0$, все корни из которого равны нулю.

Например, так как $2i=2\left(\cos\frac{\pi}{2}+i\sin\frac{\pi}{2}\right)$, то

$$(\sqrt{2i})_{1,\,2}=(\sqrt{2})_{об}\left(\cos\frac{\frac{\pi}{2}+2k\pi}{2}+i\sin\frac{\frac{\pi}{2}+2k\pi}{2}\right)\qquad(k=0,\ 1),$$

откуда легко сосчитать $(\sqrt{2i})_1=1+i$, $(\sqrt{2i})_2=-1-i$.

Другой пример: так как $1=1(\cos 0+i\sin 0)$, то

$$(\sqrt[3]{1})_{1,\,2,\,3}=(\sqrt[3]{1})_{об}\left(\cos\frac{2k\pi}{3}+i\sin\frac{2k\pi}{3}\right)\qquad(k=0,\ 1,\ 2),$$

откуда $(\sqrt[3]{1})_1=1$, $(\sqrt[3]{1})_2=-\frac{1}{2}+i\frac{\sqrt{3}}{2}$, $(\sqrt[3]{1})_3=-\frac{1}{2}-i\frac{\sqrt{3}}{2}$; в данном случае один из корней получился обычный, вещественный, а два других — мнимые.

Геометрический смысл формулы (4) показан на рис. 185, где принято $n=5$.

3. Сопряженные комплексные числа. *Сопряженным к числу $z=x+iy$ называется число $z^*=x-iy$*; часто вместо z^* пишут также $\bar{z}$. Простые свойства сопряженных чисел таковы:

1) $(z^*)^*=(x-iy)^*=[x+i(-y)]^*=x-i(-y)=x+iy=z$, т. е. числа z и z^* являются взаимно сопряженными;

2) $z+z^*=2\operatorname{Re}z$, $z-z^*=2i\operatorname{Im}z$;

3) $z^* = z$ в том и только том случае, если z вещественное;

4) $zz^* = (x - iy)(x + iy) = x^2 + y^2 = |z|^2$;

5) $|z^*| = |z|$, $\operatorname{Arg} z^* = -\operatorname{Arg} z$, т. е. точки z и z^* симметричны относительно вещественной оси;

6) $(z_1 + z_2)^* = \overset{*}{z_1} + \overset{*}{z_2}$, так как

$$(z_1 + z_2)^* = (x_1 + iy_1 + x_2 + iy_2)^* = [x_1 + x_2 + i(y_1 + y_2)]^* =$$
$$= x_1 + x_2 - i(y_1 + y_2) = (x_1 - iy_1) + (x_2 - iy_2) = \overset{*}{z_1} + \overset{*}{z_2};$$

7) $(z_1 z_2)^* = \overset{*}{z_1}\overset{*}{z_2}$, что проверяется аналогично свойству 6).

Если в свойстве 7) вместо z_1 подставить $\frac{z_1}{z_2}$, получим $\overset{*}{z_1} = \left(\frac{z_1}{z_2}\right)^* \overset{*}{z_2}$, откуда

8) $\left(\frac{z_1}{z_2}\right)^* = \frac{\overset{*}{z_1}}{\overset{*}{z_2}}$.

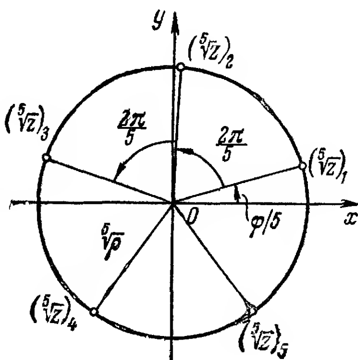


Рис. 185.

Свойства 6) и 7) автоматически распространяются на любое число слагаемых или сомножителей. Отсюда

$$(z^n)^* = (z^*)^n;$$
$$(2z^n + iz^m)^* = (2z^n)^* + (iz^m)^* =$$
$$= 2(z^*)^n - i(z^*)^m \text{ и т. д.}$$

Вообще, чтобы перейти от любого рационального выражения, содержащего любое количество переменных и коэффициентов, к сопряженному выражению, нужно каждую переменную и каждый коэффициент заменить на сопряженную величину. Можно показать, что это правило справедливо не только для рациональных выражений, но и для иррациональных, для сумм степенных рядов и т. д. Отсюда следует, что любое равенство между комплексными выражениями описанного вида остается справедливым, если в этом равенстве всюду i заменить на $-i$, так как при этом мы перейдем от равенства комплексных чисел к равенству сопряженных чисел. Поэтому *числа i и $-i$ алгебраически неразличимы*; в частности, ошибочным является распространенное мнение, что $i = \sqrt{-1}$, $-i = -\sqrt{-1}$, на самом деле $\sqrt{-1}$ имеет два значения: $\pm i$.

Сопряженные числа применяются, в частности, если в дроби вида $\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2}$ надо отделить вещественную часть от мнимой. Для этого умножают числитель и знаменатель на $\overset{*}{z_2}$, после чего знаменатель становится вещественным и требуемое отделение легко осуществить.

Например,

$$\operatorname{Re}\frac{2 + i5}{3 - i2} = \operatorname{Re}\frac{(2 + i5)(3 + i2)}{(3 - i2)(3 + i2)} = \operatorname{Re}\frac{6 + i4 + i15 - 10}{13} = \operatorname{Re}\left(-\frac{4}{13} + i\,\frac{19}{13}\right) = -\frac{4}{13}.$$

4. Формула Эйлера. Перейдем к трансцендентным действиям над комплексными числами. В п. IV.16 мы видели, что для вещественных x

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \ldots \tag{5}$$

Если вместо x поставить z, получим *определение экспоненты с комплексным показателем*: по определению

$$e^z = 1 + \frac{1}{1!}z + \frac{1}{2!}z^2 + \frac{1}{3!}z^3 + \ldots \tag{6}$$

В п. XVII.14 мы увидим, что это определение имеет смысл для всех z, а также что при этом сохраняется основное свойство экспоненты:

$$e^{z_1}e^{z_2} = e^{z_1+z_2}. \tag{7}$$

Формулы (6) и (5) показывают, что в частном случае, когда z вещественно, это новое определение e^z совпадает со старым; вообще *всякое новое определение не должно противоречить уже установленным фактам*. В то же время формула (7) подтверждает целесообразность именно данного определения e^z.

Подобным образом можно стандартным способом определить для комплексных значений аргумента функции, первоначально заданные лишь для вещественных значений аргумента. Для этого надо заданную функцию $f(x)$ разложить в ряд Тейлора (п. IV.16) по степеням x или степеням $x-a$, где a — какое-либо вещественное число, после чего заменить x на z и обозначить сумму полученного ряда через $f(z)$. Так, подобно (6), получаем на основе формул (IV.56) и (IV.57) для комплексных z

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \ldots, \tag{8}$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \ldots \tag{9}$$

и т. д. При этом, как мы увидим в п. XVII.14, все основные формулы, имеющие характер тождественных равенств и справедливые для вещественных значений аргумента (такие, например, как $\sin(-x) \equiv -\sin x$, $\sin^2 x + \cos^2 x \equiv 1$ и т. п.), остаются в силе и для комплексных его значений.

На основе приведенных формул вскрывается глубокая *связь показательной функции с тригонометрическими*. Именно, если в (6) подставить iz вместо z, мы получим

$$e^{iz} = \left(1 + \frac{1}{1!}iz - \frac{1}{2!}z^2 - \frac{1}{3!}iz^3 + \frac{1}{4!}z^4 + \frac{1}{5!}iz^5 - \frac{1}{6!}z^6 - \right.$$
$$\left. - \frac{1}{7!}iz^7 + \ldots\right) = \left(1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \frac{1}{6!}z^6 + \ldots\right) +$$
$$+ i\left(\frac{1}{1!}z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \frac{1}{7!}z^7 + \ldots\right).$$

Отсюда с помощью (8) и (9) выводим очень важную ***формулу Эйлера***

$$e^{iz}=\cos z+i\sin z. \qquad (10)$$

Применяются также формула

$$e^{-iz}=e^{i(-z)}=\cos(-z)+i\sin(-z)=\cos z-i\sin z$$

и вытекающие из нее и (10) формулы

$$\cos z=\frac{e^{iz}+e^{-iz}}{2}, \qquad \sin z=\frac{e^{iz}-e^{-iz}}{2i}. \qquad (11)$$

Все эти формулы были найдены Эйлером в 1743 г.

Из формулы Эйлера (10) на основе свойства (7) получаем выражение для экспоненты с любым комплексным показателем

$$e^{z}=e^{x+iy}=e^{x}e^{iy}=e^{x}(\cos y+i\sin y). \qquad (12)$$

Сравнение с тригонометрической формой (3) показывает, что

$$|e^{z}|=e^{x}, \quad \operatorname{Arg} e^{z}=y+2k\pi. \qquad (13)$$

В частности, видно, что всегда $|e^{z}|>0$, т. е. $e^{z}\neq 0$. Если в формуле (12) вместо e^{z} писать z, то на основании (13) получим

$$z=|z|(\cos\arg z+i\sin\arg z)=|z|e^{i\arg z}=\rho e^{i\varphi}.$$

Такая «показательная форма» комплексного числа бывает удобна для выполнения над ним алгебраических действий.

Из формул (11) вытекают соотношения между тригонометрическими и гиперболическими (п. 1.28) функциями: $\cos z=\operatorname{ch} iz$, $\sin z=\frac{\operatorname{sh} iz}{i}$, т. е. $\operatorname{sh} iz=i\sin z$. Отсюда, подставив iz вместо z, получим также $\cos iz=\operatorname{ch} z$, $\sin iz=i\operatorname{sh} z$.

В этом состоит причина глубокой связи между указанными функциями, о которой мы говорили в п. 1.28 и на основе которой можно переходить от соотношений между тригонометрическими функциями к соотношениям для гиперболических функций и обратно. (Получите основную связь между $\operatorname{ch} z$ и $\operatorname{sh} z$, подставив iz вместо z в формулу $\cos^2 z+\sin^2 z=1$.)

С помощью формул (11) легко получить также выражения степеней синуса и косинуса через тригонометрические функции кратных аргументов; например,

$$\cos^3 x=\left(\frac{e^{ix}+e^{-ix}}{2}\right)^3=\frac{e^{i\cdot 3x}+3e^{ix}+3e^{-ix}+e^{-i\cdot 3x}}{8}=$$

$$=\frac{e^{i\cdot 3x}+e^{-i\cdot 3x}}{8}+\frac{3}{8}(e^{ix}+e^{-ix})=\frac{\cos 3x}{4}+\frac{3\cos x}{4} \qquad (14)$$

и т. п. Такое преобразование применяется при интегрировании.

5. Логарифмы комплексных чисел. Определение *«комплексных логарифмов»* дается такое же, как вещественных: логарифм числа z — это число w, для которого $z=e^w$. Чтобы найти, чему равен такой логарифм, обозначим

$$z=\rho(\cos\varphi+i\sin\varphi), \qquad w=u+iv.$$

Тогда из формулы (12) получаем

$$\rho(\cos\varphi+i\sin\varphi)=z=e^w=e^u(\cos v+i\sin v).$$

Отсюда, так как u и v вещественные,

$$e^u=\rho, \quad \text{т. е.} \quad u=\ln\rho, \qquad v=\varphi+2k\pi \qquad (k \text{ — целое}),$$

где под $\ln\rho$ понимается «обыкновенный», вещественный логарифм положительного числа. Итак,

$$\operatorname{Ln} z=w=u+iv=\ln\rho+i\varphi+i2k\pi=\ln|z|+i\operatorname{Arg} z,$$

где под Ln понимается совокупность *всех* значений логарифма. Итак, *логарифм комплексного числа имеет бесконечное количество различных значений*. Единственным исключением является число «нуль», которое не имеет логарифма; можно условно написать, что $\operatorname{Ln} 0=-\infty+iv$, где v — произвольное.

Так как вещественные положительные числа — это частный случай комплексных, то и их логарифм имеет бесконечное число значений, из них одно — «обыкновенное», вещественное, а остальные — мнимые. Например,

$$\operatorname{Ln} 1=\ln 1+i0+i2k\pi=i2k\pi \qquad (k=0, \pm 1, \pm 2, \ldots);$$

при $k=0$ получаем старое значение $\ln 1=0$, но также за логарифм 1 можно принять $i2\pi$, $-i2\pi$, $i4\pi$ и т. д. Проверим это еще раз:

$$e^{i2k\pi}=\cos 2k\pi+i\sin 2k\pi=1+i0=1 \qquad (k=0, \pm 1, \pm 2, \ldots). \tag{15}$$

Отрицательные числа также имеют логарифмы, однако все их значения мнимые. Например, $\operatorname{Ln}(-1)=i\pi(2k+1)$ (проверьте!).

С помощью логарифмов определяется *возведение комплексного числа в любую комплексную степень:*

$$z_1^{z_2}=(e^{\operatorname{Ln} z_1})^{z_2}=e^{z_2\operatorname{Ln} z_1},$$

причем правая часть вычисляется по формуле (12). Так как логарифм имеет бесконечное число значений, то и вся степень имеет, вообще говоря, бесконечное число значений.

§ 2. Комплексные функции от вещественного аргумента

6. Определение и свойства. Иногда приходится рассматривать такие функции, для которых хотя *независимая переменная вещественна, но сама функция принимает комплексные значения*. Примерами могут служить

$$1)\ z=(t+i)^3; \quad 2)\ z=Me^{pt} \qquad (p=a+i\omega) \text{ и т. д.};$$

здесь независимая переменная обозначена буквой t, а функция — буквой z. Если разложить значение функции на вещественную и мнимую части, $z=x+iy$, то каждая из этих частей будет функцией t; так, в приведенных примерах получим (проверьте, считая M вещественным!)

$$1)\ x=t^3-3t,\ y=3t^2-1; \quad 2)\ x=Me^{at}\cos\omega t,\ y=Me^{at}\sin\omega t.$$

В общем случае, если

$$z = f(t) = \varphi(t) + i\psi(t), \tag{16}$$

то мы получаем

$$x = \varphi(t), \quad y = \psi(t); \tag{17}$$

обратно, от (17) можно перейти к (16). Таким образом, задание комплексной функции от вещественного аргумента равносильно заданию двух обычных, вещественных функций от того же аргумента. Эта картина совершенно аналогична заданию векторной функции от скалярного аргумента (п. VII.23); аналогия станет еще большей, если истолковывать комплексные числа как векторы (п. 2).

В результате теория комплексных функций от вещественного аргумента не имеет существенно новых черт по сравнению с теорией вещественных функций. В частности, определения непрерывности, производной и т. п. переносятся без изменений. При этом, как вытекает из рассмотрений п. XVII.14, все формулы дифференцирования сохраняются, например,

$$[(t+i)^3]' = 3(t+i)^2, \quad (Me^{pt})' = Mpe^{pt} \text{ и т. п.}$$

Изображается функция (16) линией в комплексной плоскости с параметрическими уравнениями (17).

При применении функций вида (16) надо иметь в виду следующие очевидные свойства:

если комплексные функции складываются, то и их вещественные части складываются, а также их мнимые части складываются;

если комплексная функция умножается на вещественную постоянную или вещественную функцию, то вещественная и мнимая части получают тот же множитель;

если комплексную функцию продифференцировать, то над ее вещественной и мнимой частями произведется то же действие.

Формулами эти свойства можно записать так:

$$\mathrm{Re}[f_1(t) + f_2(t)] = \mathrm{Re}\, f_1(t) + \mathrm{Re}\, f_2(t) \text{ и т. д. (проделайте это!).}$$

Эти свойства дают возможность, вместо того чтобы производить указанные действия над вещественной или мнимой частью, осуществить эти действия над всей комплексной функцией, а от результата взять вещественную или соответственно мнимую часть. Замечательно, что такой переход к комплексным величинам с обратным переходом к искомым вещественным величинам может оказаться проще и нагляднее, чем непосредственные действия над вещественными величинами.

Применяется также функция

$$\mathrm{Ln}[t-(a+ib)] = \ln|t-a-ib| + i\,\mathrm{Arg}[t-a-ib] = \frac{1}{2}\ln[(t-a)^2+b^2] + i\left[\mathrm{arc\,tg}\,\frac{t-a}{b} - \frac{\pi}{2} + k\pi\right],$$

где значение целого числа k еще надо выбрать.

7. Применение к описанию колебаний. Функцию

$$U(t)=Me^{i(\omega t+\alpha)}=M\cos(\omega t+\alpha)+iM\sin(\omega t+\alpha)\quad (M>0,\ \omega>0) \tag{18}$$

удобно применить для исследования гармонических колебаний (ср. п. I.29). Для этого надо заметить, что величина (18) имеет модуль M и аргумент $\omega t+\alpha$, т. е. она представима *вектором постоянной длины, который равномерно вращается с угловой скоростью* ω.

Рассмотрим, например, наложение колебаний, происходящих с одинаковой частотой. Пусть надо сложить две величины: $u_1(t)=M_1\sin(\omega t+\alpha_1)$ и $u_2(t)=M_2\sin(\omega t+\alpha_2)$. Для этого введем соответствующие комплексные величины $U_1(t)=M_1e^{i(\omega t+\alpha_1)}$ и $U_2(t)=M_2e^{i(\omega t+\alpha_2)}$, у которых u_1 и u_2 будут служить мнимыми частями. Векторы $U_1(t)$ и $U_2(t)$ равномерно вращаются с угловой скоростью ω; значит, и вектор $U_1(t)+U_2(t)$ равномерно вращается с той же скоростью и потому его можно представить в виде (18). Чтобы найти M и α, достаточно рассмотреть картину в момент $t=0$ (рис. 186); из нее, проектируя на оси координат, получим

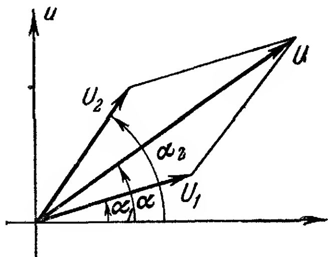


Рис. 186.

$$\begin{aligned} M\cos\alpha&=M_1\cos\alpha_1+M_2\cos\alpha_2,\\ M\sin\alpha&=M_1\sin\alpha_1+M_2\sin\alpha_2.\end{aligned} \tag{19}$$

Взяв мнимую часть от $U(t)$, мы приходим окончательно к выводу, что $u_1(t)+u_2(t)=M\sin(\omega t+\alpha)$, где M и α находятся из равенств (19) (найдите их и объясните формулу для M с помощью рис. 186).

Аналогичный результат получается при наложении любого числа гармонических колебаний, происходящих с одинаковой частотой. О наложении колебаний различной частоты будет сказано в п. XVII.23.

Преимущество экспоненты (18) перед тригонометрическими функциями особенно проявляется при дифференцировании:

$$\frac{dU}{dt}=Me^{i(\omega t+\alpha)}i\omega=i\omega U. \tag{20}$$

В силу п. 2 получается вектор, также равномерно вращающийся с угловой скоростью ω, но *опережающий* U *на* 90° *и имеющий модуль, в* ω *раз больший.* При дальнейших дифференцированиях эти поворот и растяжение повторяются.

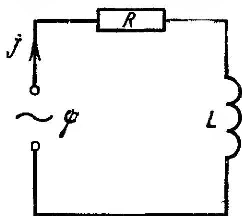


Рис. 187.

Покажем применение функций вида (18) к расчету установившегося тока в цепи с активным сопротивлением и индуктивностью (рис. 187), если к цепи подключен источник напряжения, меняющегося по гармоническому закону $\varphi=\varphi_0\sin(\omega t+\beta)$. При этом в цепи возникает ток, также меняющийся по гармоническому закону $j=j_0\sin(\omega t+\alpha)$, однако j_0 и α нам заранее не известны. Приравнивая φ сумме падений напряжений на R и L на основе известных в физике законов, получаем основное уравнение задачи $Rj+L\dfrac{dj}{dt}=\varphi$.

Введем понятия *комплексного напряжения* и *комплексной силы тока* по формулам $\Phi=\varphi_0e^{i(\omega t+\beta)}$, $J=j_0e^{i(\omega t+\alpha)}$; «настоящие» напряжение и сила тока служат мнимыми частями этих выражений. В силу свойств, описанных в п. 6, для получения j надо решить уравнение $RJ+L\dfrac{dJ}{dt}=\Phi$, а от ответа взять

мнимую часть. Согласно формуле (20) получаем

$$RJ+Li\omega J=\Phi, \quad \text{т. е.} \quad J=\frac{\Phi}{R+i\omega L}. \tag{21}$$

Мы видим, что индуктивность L можно истолковать как некое сопротивление, численно равное $i\omega L$; это значение называется *кажущимся сопротивлением* или *импедансом* элемента L. Записав $(R+i\omega L)^{-1}$ в показательной форме $re^{-i\vartheta}$, из (21) получим (проверьте!)

$$j_0=r\varphi_0=\varphi_0(R^2+\omega^2L^2)^{-1/2}, \quad \alpha=\beta-\operatorname{arctg}\frac{\omega L}{R}. \tag{22}$$

(Попробуйте получить вектор $j_0e^{i\alpha}=J\,|_{t=0}$ на основе первого равенства (21) с помощью геометрического построения, из которого и вывести формулы (22).)

§ 3. Понятие о функциях комплексного переменного

Теория и приложения комплексных функций от комплексного аргумента содержат много принципиально нового по сравнению с функциями вещественного аргумента. Эта теория составляет содержание многих полезных книг, из которых для первоначального ознакомления можно избрать [39], [43, т. 3, ч. 1], [49]. Мы приведем здесь лишь несколько простых фактов, непосредственно примыкающих к остальному материалу книги.

8. Разложение многочлена на множители. Рассмотрим целую рациональную функцию от $z=x+iy$, т. е. многочлен

$$P(z)=a_0z^n+a_1z^{n-1}+\ldots+a_{n-1}z+a_n \qquad (a_0\neq 0) \tag{23}$$

степени n с, вообще говоря, комплексными коэффициентами. В указанных выше книгах приводится доказательство замечательной «*основной теоремы алгебры*»; эта теорема гласит, что *всякий многочлен степени* $n\geqslant 1$ *имеет по крайней мере один комплексный* (вещественный или мнимый) *нуль*, т. е. корень уравнения

$$P(z)=0. \tag{24}$$

(Эта теорема доказана Даламбером в середине XVIII в. и — с большей точностью — Гауссом в самом конце XVIII в.) Если обозначить через z_1 какой-либо из этих корней, то, как доказывается в элементарной алгебре, $P(z)$ нацело делится на двучлен $z-z_1$, т. е. $P(z)=(z-z_1)P_1(z)$, где $P_1(z)$ — уже многочлен $(n-1)$-й степени. Если к нему опять применить те же рассуждения, то мы получим $P_1(z)=(z-z_2)P_2(z)$, т. е. $P(z)=(z-z_1)(z-z_2)P_2(z)$, где $P_2(z)$ — многочлен $(n-2)$-й степени. Эти рассуждения можно продолжать, пока мы не дойдем до «многочлена нулевой степени», т. е. постоянной, и получим

$$P(z)=a_0(z-z_1)(z-z_2)\ \ldots\ (z-z_n). \tag{25}$$

Из этой формулы видно, что все числа $z_1, z_2, \ldots, z_n$ являются нулями многочлена $P(z)$ и никаких других нулей этот многочлен

не имеет. *Итак, алгебраическое уравнение* (24) *степени n имеет равно n корней.*

Некоторые из корней уравнения (24) могут совпадать друг с другом, повторяться; такие корни называются *кратными* (двойными, т. е. кратности 2, тройными, т. е. кратности 3, и т. д.) в отличие от *простых*, т. е. неповторяющихся корней. Конечно, при подсчете числа корней каждый из них должен браться со своей кратностью.

Простой признак кратности корня вытекает из формулы Тейлора для многочленов (IV.46), которая, как мы поясним в п. 11, справедлива и для многочленов от комплексного переменного. Пусть $z=a$ — корень уравнения (24). Тогда $P(a)=0$ и из указанной формулы вытекает, что если

$$P(a)=P'(a)=\ldots=P^{(k-1)}(a)=0; \quad P^{(k)}(a)\neq 0, \tag{26}$$

то многочлен $P(z)$ делится на $z-a$ точно в k-й степени, т. е. значение $z=a$ является корнем уравнения (24) кратности k.

Если $P(z)$ представляет собой произвольную, хотя бы даже трансцендентную функцию и для некоторого значения $z=a$ выполняются соотношения (26), то значение $z=a$ *называется* корнем уравнения (24) *кратности* k. (В частности, если $P(a)=0$, $P'(a)\neq 0$, то значение $z=a$ будет *простым* корнем.) В этом случае из ряда Тейлора (IV.53) видим, что отношение $P(z)/(z-a)^k$ имеет при $z\to a$ конечный предел, не равный нулю, т. е. имеет при $z=a$ устранимый разрыв (см. п. III.13). Поэтому можно написать, что

$$P(z)=(z-a)^k\frac{P(z)}{(z-a)^k}=(z-a)^k Q(z),$$

где функция $Q(z)$ остается непрерывной при $z=a$, причем $Q(a)\neq 0$.

Например, значение $x=0$ является тройным корнем уравнения

$$x-\sin x=0,$$

так как $(x-\sin x)|_{x=0}=0$; $(1-\cos x)|_{x=0}=0$; $(\sin x)|_{x=0}=0$; $(\cos x)|_{x=0}=1$.

Если в разложении (25) объединить равные множители, то получим

$$P(z)=a_0(z-z_1)^{\alpha_1}\ldots(z-z_k)^{\alpha_k}, \tag{27}$$

где $z_1,\ldots,z_k$ — все *различные* корни уравнения (24), а $\alpha_1,\ldots,\alpha_k$ — их кратности. Разложения (25) и (27) справедливы для многочленов (23) как с вещественными, так и с комплексными коэффициентами.

Если многочлен (23) *имеет вещественные коэффициенты, то уравнение* (24) *всегда вместе с мнимым корнем обладает сопряженным корнем, причем той же кратности.* Действительно, если $P(z_m)=0$, то $[P(z_m)]^*=0^*=0$. Но в силу п. 3

$$[P(z_m)]^*=[a_0z_m^n+\ldots+a_n]^*=a_0^*(z_m^*)^n+\ldots+a_n^*=$$
$$=a_0(z_m^*)^n+\ldots+a_n=P(z_m^*),$$

откуда и следует, что $z=z_m^*$ будет решением уравнения (24). Если же $z=z_m$ будет двойным корнем уравнения (24), то будет дополнительно $P'(z_m)=0$, откуда аналогично получаем, что $P'(z_m^*)=0$, и т. д.

Если в разложении (27) объединить множители, отвечающие сопряженным корням, то получим

$$[z-(\alpha+i\beta)]\,[z-(\alpha-i\beta)]=(z-\alpha)^2+\beta^2=$$
$$=z^2+pz+q \quad (p=-2\alpha,\ q=\alpha^2+\beta^2). \qquad (28)$$

Такое объединение применяется при разложении многочлена с вещественными коэффициентами от вещественного аргумента, который естественно обозначать буквой x; тогда из (27) получаем

$$P(x)=a_0(x-x_1)^{\alpha_1}\ldots(x-x_r)^{\alpha_r}(x^2+p_1x+q_1)^{\beta_1}\ldots$$
$$\ldots(x^2+p_sx+q_s)^{\beta_s}, \qquad (29)$$

где первые r скобок отвечают вещественным корням, а последние s — парам сопряженных мнимых корней. Так как p и q вещественны, то мы получаем, что *всякий вещественный многочлен* (т. е. многочлен с вещественными коэффициентами) *можно разложить на вещественные линейные и квадратичные множители.* Конечно, если все корни вещественные, то в разложении (29) присутствуют только линейные множители, а если все корни мнимые, то присутствуют только квадратичные множители. Показатели $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s$ равны кратностям соответствующих корней; в частности, для простых корней они равны 1.

9. Численное решение алгебраических уравнений. Чтобы фактически осуществить разложение (27) или (29), нужно решить уравнение (24), т. е. алгебраическое уравнение n-й степени. Решение при $n=2$ хорошо известно из школьного курса. В учебниках высшей алгебры можно найти формулы для решения при $n=3$ и $n=4$, найденные уже в XVI в.; однако эти формулы столь сложны, особенно при $n=4$, что на практике почти не применяются. Если же $n>4$, то общих формул для решения, выражающих решение через коэффициенты уравнения с помощью алгебраических действий, не существует. Это доказали Абель и заложивший основы современной алгебры французский математик Э. Галуá (1811—1832).

Однако уравнения можно решать приближенно с любой требуемой степенью точности! В § V.1 были описаны методы нахождения вещественных корней любых уравнений вида $f(x)=0$. Для нахождения мнимых корней можно воспользоваться методом Ньютона (см. формулу (V.7), где x надо считать мнимым) или методом итераций (п. V.3) в любой форме; можно также подставить в уравнение (24) $z=x+iy$, после чего отделить вещественную часть от мнимой:

$$P(z)=P(x+iy)=Q(x,\,y)+iR(x,\,y).$$

Тогда уравнение (24) сведется к системе уравнений

$$\left.\begin{aligned}Q(x,\,y)=0,\\ R(x,\,y)=0,\end{aligned}\right\}$$

которую можно решать, как будет описано в п. XII.12. Эти методы описаны в книгах [3], [8], [17], [35], [41].

Имеются методы, пригодные только для алгебраических уравнений. Мы опишем метод, предложенный несколькими авторами, в том числе в 1834 г. Н. И. Лобачевским.

Любое алгебраическое уравнение можно записать в виде

$$z^n+a_1z^{n-1}+a_2z^{n-2}+\ldots+a_n=0 \qquad (30)$$

с коэффициентом при старшей степени, равным 1 (почему?). Если отделить четные степени от нечетных

$$z^n+a_2z^{n-2}+\ldots=-a_1z^{n-1}-a_3z^{n-3}-\ldots,$$

а затем обе части этого равенства возвысить в квадрат, получим уравнение, содержащее только четные степени z. Поэтому если обозначить $z^2=p$, то мы получим для p уравнение степени n (почему?), корнями которого служат квадраты корней уравнения (30). Если над полученным уравнением совершить такое же преобразование, а затем обозначить $p^2=q$, то мы придем к уравнению степени n, корнями которого служат четвертые степени корней уравнения (30), и т. д.

После нескольких таких преобразований наибольшее значение приобретает тот корень или те корни, модуль которых наиболее велик; например, если уравнение (30) имеет корни $z_1=2$, $z_2=-1$, $z_3=\frac{1}{2}$, то следующее уравнение имеет корни $p_1=4$, $p_2=1$, $p_3=\frac{1}{4}$, следующее — корни $q_1=16$, $q_2=1$, $q_3=\frac{1}{16}$ и т. д. Если произведено m преобразований, то мы придем к уравнению вида

$$v^n+C_1v^{n-1}+\ldots+C_n=0, \tag{31}$$

корни которого, пока еще нам неизвестные, равны $v_1=z_1^{2^m}$, ..., $v_n=z_n^{2^m}$. Поэтому на основе формулы (25) уравнение (31) имеет вид

$$(v-z_1^{2^m})\ldots(v-z_n^{2^m})=0.$$

Сделаем в уравнении (31) замену $v=10^l\bar{v}$, подобрав целое число l так, чтобы в получающемся уравнении для $\bar{v}$ после деления на старший коэффициент 10^{nl} наибольший из модулей коэффициентов при $\bar{v}^{n-1}$, $\bar{v}^{n-2}$, ..., 1 имел порядок 1, т. е. не был ни слишком велик, ни слишком мал. Так как полученное уравнение после разложения левой части должно иметь вид

$$(\bar{v}-\bar{v}_1)\ldots(\bar{v}-\bar{v}_n)=0 \qquad (\text{где } \bar{v}_1=10^{-l}v_1=10^{-l}z_1^{2^m} \text{ и т. д.}), \tag{32}$$

то заключаем, что старший (или старшие) из корней уравнения (32) имеет порядок 1, тогда как остальные пренебрежимо малы. Отбрасывая эти корни, т. е. полагая их равными нулю, мы тем самым отбрасываем в уравнении для $\bar{v}$ члены со слишком малыми коэффициентами.

Решив уравнение для $\bar{v}$, полученное после указанного отбрасывания, мы найдем приближенные значения наибольших по модулю корней $\bar{v}$, после чего, возвращаясь к z, мы получим приближенно наибольшие по модулю корни уравнения (30), причем тем точнее, чем больше m. Некоторая трудность при переходе от $\bar{v}$ к z возникает при извлечении корня степени 2^m. Если уравнение (30) имеет вещественные коэффициенты, а старший по модулю корень уравнения для $\bar{v}$ получится только один, то и старший по модулю корень уравнения (30) имеется только один и притом вещественный (почему?); значит, при извлечении корня надо ограничиться лишь двумя вещественными значениями. Если же указанные условия не выполнены, то надо извлекать корень по правилам п. 2, и получится много возможных значений. Чтобы узнать, какое именно значение нужно взять, можно все эти значения подставить в уравнение (30) и проверить, при каком оно лучше удовлетворяется. Можно также воспользоваться *соотношениями между корнями алгебраического уравнения и его коэффициентами*.

Чтобы вывести эти соотношения, надо сравнить коэффициенты при одинаковых степенях z в формулах (23) и (25), для чего в последней раскрыть скобки; получится

$$\left.\begin{aligned} &z_1+z_2+\ldots+z_n=-\frac{a_1}{a_2},\\ &z_1z_2+z_1z_3+\ldots+z_{n-1}z_n=\frac{a_2}{a_0}\\ &\text{(всевозможные произведения по два)},\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ &z_1z_2\ldots z_n=(-1)^n\frac{a_n}{a_0}. \end{aligned}\right\} \quad (33)$$

Найдя корень v_1 уравнения (31), можно исключить при помощи деления двучлен $v-v_1$ из этого уравнения, после чего аналогичным путем найти следующий корень, и т. д. Часто эти корни можно найти с помощью соотношений (33), примененных к уравнению (31); при этом надо иметь в виду, что если корни уравнения (31) резко различаются по модулю, то в левых частях соотношений (33) можно отбрасывать слагаемые.

Найдя корни уравнения (30) приближенно, можно для их уточнения применить один из итерационных методов, описанных в § V.1, например метод Ньютона (п. V.2).

Рассмотрим, например, уравнение $z^3+z^2-3=0$, уже решенное в п. V.2. Последовательное преобразование по Лобачевскому дает (проверьте!):

$$p^3-p^2+6p-9=0, \quad q^3+11q^2+18q-81=0,$$
$$u^3-85u^2+2106u-6561=0.$$

Остановимся на этом и сделаем подстановку $u=10^l\bar{u}$, получим

$$\bar{u}^3-\frac{85}{10^l}\bar{u}^2+\frac{2106}{10^{2l}}\bar{u}-\frac{6561}{10^{3l}}=0. \quad (34)$$

В данном случае нужно взять $l=2$ (проверьте, выбрав другие значения l). После этого свободный член в уравнении (34) получится малым и его можно отбросить, что даст

$$\bar{u}^2-0{,}85\bar{u}+0{,}2106=0. \quad (35)$$

Таким образом, $\bar{u}_1$ и $\bar{u}_2$, а потому z_1 и z_2 — мнимые сопряженные. Далее, так как $\bar{u}_1\bar{u}_2=0{,}2106$, то $u_1u_2=2106$ и в силу равенства $(z_1z_2)^8=u_1u_2$ получаем

$$z_1z_2=|z_{1,2}|^2=\sqrt[8]{2106}=2{,}60.$$

Но $z_1z_2z_3=3$ (почему?), т. е. $z_3=\dfrac{3}{2{,}60}=1{,}15$.

Кроме того, $z_1+z_2+z_3=-1$ (почему?); значит, если обозначить $z_{1,2}=\alpha\pm i\beta$, то $2\alpha+1{,}15=-1$, откуда $\alpha=-1{,}08$. Но $\alpha^2+\beta^2=|z_{1,2}|^2=2{,}60$, т. е. $\beta=\sqrt{2{,}60-1{,}08^2}=1{,}20$. Итак, приближенно $z_{1,2}=-1{,}08\pm i1{,}20$, $z_3=1{,}15$.

Итерация по формуле Ньютона (V.6) дает более точные значения $z_{1,2}=-1{,}087\pm i1{,}172$, $z_3=1{,}175$.

Метод Лобачевского особенно удобен, если у заданного уравнения все корни вещественные и различные по модулю. В некоторых случаях возможна рационализация вычислений; так, в разобранном примере можно было

вычислять только z_3, после чего произвести деление на $z-z_3$ и перейти к квадратному уравнению.

Метод Лобачевского и сходные методы подробно разобраны в книгах [20], [28], [33], [50].

10. Разложение дробно-рациональной функции на простейшие рациональные дроби. Напомним (п. I.17), что дробно-рациональная функция — это отношение двух многочленов

$$f(z)=\frac{Q(z)}{P(z)}=\frac{b_0z^m+\ldots+b_m}{a_0z^n+\ldots+a_n}. \tag{36}$$

Если $m<n$, то дробь называется *правильной*, независимо от значений коэффициентов, в противном случае — *неправильной*. *Неправильную дробь всегда можно представить в виде суммы целой части* (многочлена) *и правильной дроби*; это можно сделать, например, посредством деления числителя на знаменатель столбиком, как в школе. Так,

$$\frac{z^3}{z^2-3}=z+\frac{3z}{z^2-3},\quad \frac{z^2-1}{2z^2+5}=\frac{1}{2}+\frac{-\frac{7}{2}}{2z^2+5} \text{ и т. д.}$$

Важно отметить, что в отличие от числовых дробей *сумма правильных дробно-рациональных функций также является правильной дробью*. Чтобы это доказать, будем указывать степень многочлена с помощью индекса; тогда

$$\frac{Q_m(z)}{P_n(z)}+\frac{\overline{Q}_{\overline{m}}(z)}{\overline{P}_{\overline{n}}(z)}=\frac{Q_m(z)\,\overline{P}_{\overline{n}}(z)+\overline{Q}_{\overline{m}}(z)\,P_n(z)}{P_n(z)\overline{P}_{\overline{n}}(z)}.$$

Если слева стоят правильные дроби, то $m<n$ и $\overline{m}<\overline{n}$. Но тогда в числителе суммы первое слагаемое имеет степень $m+\overline{n}<n+\overline{n}$, а второе — степень $\overline{m}+n<n+\overline{n}$, т. е. и весь числитель имеет степень $<n+\overline{n}$, а степень знаменателя $=n+\overline{n}$, т. е. сумма — правильная дробь. Это, очевидно, верно и для большего числа слагаемых.

Пусть дробь (36) правильная, причем знаменатель разложен на линейные множители и одинаковые множители объединены (см. формулу (27)). Тогда можно доказать, что эту дробь возможно представить в виде

$$\begin{aligned}\frac{Q(z)}{P(z)}=\frac{A_1}{(z-z_1)^{\alpha_1}}+\frac{A_2}{(z-z_1)^{\alpha_1-1}}+\ldots+\frac{A_{\alpha_1}}{z-z_1}+\\ +\frac{B_1}{(z-z_2)^{\alpha_2}}+\frac{B_2}{(z-z_2)^{\alpha_2-1}}+\ldots+\frac{B_{\alpha_2}}{z-z_2}+\ldots\\ \ldots+\frac{D_1}{(z-z_k)^{\alpha_k}}+\frac{D_2}{(z-z_k)^{\alpha_k-1}}+\ldots+\frac{D_{\alpha_k}}{z-z_k},\end{aligned} \tag{37}$$

где $A_1, A_2, \ldots, D_{\alpha_k}$ — некоторые числовые коэффициенты. Дроби, стоящие в правой части, называются *простейшими рациональными*

дробями первого типа. Итак, *всякую правильную рациональную дробь можно представить в виде суммы простейших рациональных дробей первого типа.*

Для доказательства этого совершим следующее преобразование:

$$\frac{Q(z)}{P(z)}=\frac{Q(z)}{a_0(z-z_1)^{\alpha_1}(z-z_2)^{\alpha_2}\ldots(z-z_k)^{\alpha_k}}=$$

$$=\frac{Q(z)\,[(z-z_1)-(z-z_2)]}{a_0(z-z_1)^{\alpha_1}(z-z_2)^{\alpha_2}\ldots(z-z_k)^{\alpha_k}(z_2-z_1)}=$$

$$=\frac{Q(z)}{a_0(z_2-z_1)(z-z_1)^{\alpha_1-1}(z-z_2)^{\alpha_2}\ldots(z-z_k)^{\alpha_k}}-$$

$$-\frac{Q(z)}{a_0(z_2-z_1)(z-z_1)^{\alpha_1}(z-z_2)^{\alpha_2-1}\ldots(z-z_k)^{\alpha_k}}.$$

В каждой из двух полученных дробей в знаменателе стало на одну скобку меньше, так как постоянный множитель z_2-z_1 можно объединить с a_0. Вновь совершая такое преобразование над каждой из этих двух дробей, мы уменьшаем число скобок в знаменателе еще на единицу и т. д. Так можно поступать, пока в знаменателе имеется по крайней мере два различных линейных множителя. Через некоторое число шагов дробей с различными множителями в знаменателе не останется, т. е. мы придем к сумме дробей вида $\frac{Q(z)}{a(z-z_l)^{\alpha}}$.

Разложив числитель по степеням $z-z_l$ (см. начало п. IV.15), получим

$$\frac{Q(z)}{a(z-z_l)^{\alpha}}=\frac{c+d(z-z_l)+\ldots+g(z-z_l)^m}{a(z-z_l)^{\alpha}}=\frac{\frac{c}{a}}{(z-z_l)^{\alpha}}+\frac{\frac{d}{a}}{(z-z_l)^{\alpha-1}}+\ldots;$$

при этом после деления может остаться и целая часть. Если произвести сложение всех полученных дробей, мы и придем к окончательной формуле (37); при этом все целые части должны при сложении взаимно уничтожиться, так как в противном случае правильная дробь была бы представлена в виде суммы правильной дроби и многочлена, чего не может быть (почему?).

На практике указанное разложение обычно осуществляют по методу неопределенных коэффициентов: выписывают правую часть формулы (37) с буквенными коэффициентами в числителях, после чего находят эти коэффициенты с помощью тех или иных манипуляций, которые мы продемонстрируем на примере.

Пусть надо разложить на простейшие дроби функцию

$$\frac{x^3-2x+3}{x(x-1)(x+2)^2}. \tag{38}$$

Так как дробь правильная, то пишем в силу формулы (37), для простоты обозначая все коэффициенты различными буквами:

$$\frac{x^3-2x+3}{x(x-1)(x+2)^2}=\frac{A}{x}+\frac{B}{x-1}+\frac{C}{(x+2)^2}+\frac{D}{x+2}. \tag{39}$$

Умножив на общий знаменатель, получим

$$x^3-2x+3=A(x-1)(x+2)^2+Bx(x+2)^2+Cx(x-1)+Dx(x-1)(x+2). \quad (40)$$

Это равенство должно быть тождеством. Поэтому раскроем скобки и приравняем коэффициенты при одинаковых степенях x; получим систему уравнений (проверьте!):

$$\begin{array}{r|l} x^3 & A+B \qquad\quad + D=1, \\ x^2 & 3A+4B+C+D=0, \\ x & \qquad 4B-C-2D=-2, \\ 1 & -4A \qquad\qquad\quad =3, \end{array} \quad (41)$$

откуда легко подсчитать

$$A=-\frac{3}{4}=-0{,}750; \quad B=\frac{2}{9}=0{,}222; \quad C=-\frac{1}{6}=-0{,}167; \quad D=\frac{55}{36}=1{,}528. \quad (42)$$

Этот способ — самый надежный (уравняв коэффициенты, мы уверены, что соотношение (40), а с ним и (39) представляют собой тождества), но не самый простой. Можно, не раскрывая скобок в равенстве (40), положить в нем x равным каким-либо четырем числам, в соответствии с числом неизвестных коэффициентов, чтобы получить уравнения для определения A, B, C, D. В данном примере лучше всего положить $x=0$, 1, -2, так как при этих значениях часть слагаемых выпадает, и еще какому угодно значению, например -1. Тогда получим соотношения

$$-4A=3; \quad 9B=2; \quad 6C=-1; \quad -2A-B+2C+2B=4, \quad (43)$$

из которых найдем те же значения (42).

Последний прием называется методом *коллока́ции*, т. е. приравнивания двух выражений при отдельных значениях x; он особенно быстро приводит к цели, если знаменатель разлагаемой дроби не имеет кратных корней, т. е. все скобки в его разложении присутствуют в первой степени.

Если дробь (36) имеет вещественные коэффициенты и независимая переменная считается вещественной, но знаменатель обладает мнимыми корнями, то разложение (37) хотя и возможно, но не всегда удобно. В этом случае часто применяется разложение другого типа. Именно, исходя из разложения (37) и применяя формулу (29), можно доказать возможность следующего разложения:

$$\frac{Q(x)}{P(x)}=\frac{Q(x)}{a_0(x-x_1)^{\alpha_1}\ldots(x-x_r)^{\alpha_r}(x^2+p_1x+q_1)^{\beta_1}\ldots(x^2+p_sx+q_s)^{\beta_s}}=$$
$$=\frac{A_1}{(x-x_1)^{\alpha_1}}+\frac{A_2}{(x-x_1)^{\alpha_1-1}}+\ldots+\frac{A_{\alpha_1}}{x-x_1}+\ldots$$
$$\ldots+\frac{D_1}{(x-x_r)^{\alpha_r}}+\frac{D_2}{(x-x_r)^{\alpha_r-1}}+\ldots+\frac{D_{\alpha_r}}{x-x_r}+$$
$$+\frac{M_1x+N_1}{(x^2+p_1x+q_1)^{\beta_1}}+\frac{M_2x+N_2}{(x^2+p_1x+q_1)^{\beta_1-1}}+\ldots$$
$$\ldots+\frac{M_{\beta_1}x+N_{\beta_1}}{x^2+p_1x+q_1}+\ldots+\frac{U_1x+V_1}{(x^2+p_sx+q_s)^{\beta_s}}+$$
$$+\frac{U_2x+V_2}{(x^2+p_sx+q_s)^{\beta_s-1}}+\ldots+\frac{U_{\beta_s}x+V_{\beta_s}}{x^2+p_sx+q_s}. \quad (44)$$

Дроби в правой части, в знаменателе которых стоит степень квадратного трехчлена, называются *простейшими рациональными дробями второго типа.* Как и раньше, все коэффициенты, стоящие в числителях, можно определить по методу неопределенных коэффициентов. Однако, так как теперь все манипуляции будут производиться лишь с вещественными числами, то неизвестные коэффициенты, определяемые из системы уравнений первой степени, полученной подобно системе (41) или (43), будут обязательно вещественными. Итак, *всякую правильную рациональную дробь с вещественными коэффициентами можно представить в виде суммы простейших рациональных дробей первого и второго типов с вещественными коэффициентами.* При этом, конечно, может получиться, что в разложении присутствуют только дроби первого типа, если все корни знаменателя вещественные, или только второго типа, если все корни знаменателя мнимые.

Мы не будем здесь доказывать возможность разложения (44). В каждом конкретном примере эта возможность подтверждается результатами вычислений.

11. Общие замечания о функциях комплексного переменного. Каждая из функций

$$w=z^3-iz,\quad w=\frac{z}{z^2+1},\quad w=ze^z \text{ и т. д.} \tag{45}$$

при комплексных z принимает комплексные значения; в общем виде можно написать $w=f(z)$, где z комплексное и w комплексное. Многие понятия из теории вещественных переменных величин без существенных изменений переносятся на комплексные переменные: это относится к понятию предела и свойствам пределов (конечно, исключая свойства, связанные с неравенствами), понятиям непрерывности и точек разрыва и т. п. Отыскание точек разрыва производится по тем же правилам, что в п. III.13 для функций вещественного переменного. Так, из функций (45) первая и третья непрерывны при всех z, а вторая имеет две точки разрыва $z=\pm i$, в которых она обращается в бесконечность.

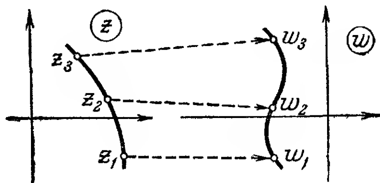


Рис. 188.
Пунктиром обозначено отображение: $w_1=f(z_1)$ и т. д.

Определение производной функции комплексного переменного также дается аналогично производной вещественной функции (п. IV.2):

$$\frac{dw}{dz}=f'(z)=\lim_{\Delta z\to 0}\frac{\Delta w}{\Delta z}=\lim_{\Delta z\to 0}\frac{f(z+\Delta z)-f(z)}{\Delta z},$$

где предел должен быть вполне определенным при произвольном стремлении Δz к нулю. Можно проверить, на чем мы не будем останавливаться, что все свойства производной и все формулы дифференцирования, выведенные в п. IV.4—5, сохраняются без изменений. Без изменения сохраняются также понятие производных высших порядков и основанные на нем формула и ряд Тейлора (п. IV.5); на этом вопросе мы еще остановимся в п. XVII.14.

Принципиально новым является геометрическое истолкование функции комплексного переменного $w=f(z)$. Так как значение z изображается точкой

в комплексной плоскости z, а соответствующее значение w — точкой в комплексной плоскости w, то мы получаем, что точкам плоскости z соответствуют точки плоскости w. Говорят также, что точки плоскости z *отображаются* на точки плоскости w, другими словами, *плоскость* z — или та ее часть, где задана функция $f(z)$, — *отображается в плоскость* w. Например, функция $w=z^3-iz$ определяет отображение, при котором точки $z=0$; 1; i; $2-i$ переходят соответственно в точки (проверьте!) $w=0$, $1-i$; $1-i$; $1-13i$ и т. п. Если точка z, двигаясь в плоскости z, описывает линию, то w также опишет некоторую линию в плоскости w (рис. 188). Итак, при отображении $w=f(z)$ линии переходят в линии; фигуры в плоскости z переходят в фигуры в плоскости w, хотя при этом переходе форма фигуры может существенно измениться.

ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

§ 1. Функции двух переменных

1. Способы задания. Понятия о функциях любого числа независимых переменных и об обозначениях для них были даны в пп. I.11 и I.12 и применялись уже в нашем курсе. Однако о способах задания таких функций надо поговорить более подробно.

Аналитический способ задания функции $z=f(x, y)$ двух переменных не отличается существенно от того, как это делается для функций одной переменной. В то же время табличный способ оформляется значительно более громоздко: так как нужно задавать значения двух независимых переменных, то приходится применять таблицу с двумя входами.

Таблица с двумя входами

$$z=f(x, y)$$

$x \backslash y$	y_1	y_2	y_3	...	y_N
x_1	$z_{11}=f(x_1, y_1)$	$z_{12}=f(x_1, y_2)$	$z_{13}=f(x_1, y_3)$	...	$z_{1N}=f(x_1, y_N)$
x_2	$z_{21}=f(x_2, y_1)$	$z_{22}=f(x_2, y_2)$	$z_{23}=f(x_2, y_3)$	...	$z_{2N}=f(x_2, y_N)$
⋮	...	...	...	...	...
x_M	$z_{M1}=f(x_M, y_1)$	$z_{M2}=f(x_M, y_2)$	$z_{M3}=f(x_M, y_3)$	...	$z_{MN}=f(x_M, y_N)$

Значения функции здесь приходится снабжать двумя индексами (двойная нумерация). Ясно, что если число значений для x и y велико, то такую таблицу трудно составить.

При составлении таблицы можно также пользоваться тем, что если фиксировать значение одной из независимых переменных, например x, то z станет функцией только y. Тогда получится система таблиц с одним входом, что, конечно, равносильно таблице с двумя входами.

Система таблиц с одним входом

$x = x_1$

y	y_1	y_2	...	y_N
z	$z_{11} = f(x_1, y_1)$	$z_{12} = f(x_1, y_2)$	...	$z_{1N} = f(x_1, y_N)$

$x = x_2$

y	y_1	y_2	...	y_N
z	$z_{21} = f(x_2, y_1)$	$z_{22} = f(x_2, y_2)$	...	$z_{2N} = f(x_2, y_N)$

и т. д.

Такой же принцип фиксирования значений одной из переменных может быть использован при графическом изображении функции двух переменных, в результате чего получится система графиков, например, такого вида, как показано на рис. 189.

В теоретических рассмотрениях встречается еще следующий способ графического изображения функции $z = f(x, y)$. Рассмотрим

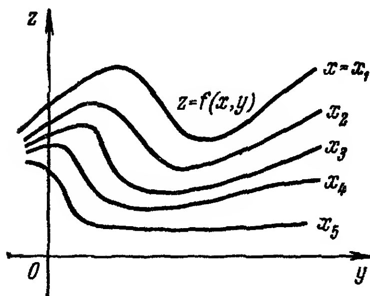


Рис. 189.

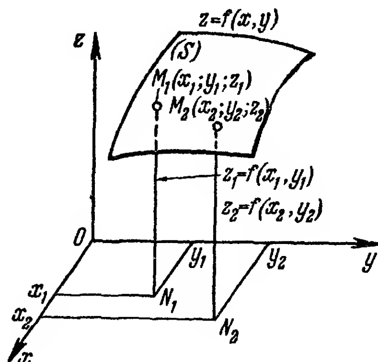


Рис. 190.

декартовы координаты x, y, z в пространстве (можно применять и другие системы координат, с которыми мы познакомимся в п. X.1). Придавая независимым переменным какие-либо численные значения $x = x_1$, $y = y_1$, получим точку N_1 (рис. 190) в *плоскости аргументов* xOy. Вычислив соответствующее значение функции $z_1 = f(x_1, y_1)$, мы сможем построить соответствующую точку M_1 в пространстве. Придавая независимым переменным другие значения, мы сможем построить точку M_2 и т. д. Если теперь теоретически представить,

что независимые переменные принимают все возможные значения, то построенные точки N заполнят всю плоскость xOy или часть ее, а так как над (или под, в зависимости от знака функции) каждой точкой N имеется соответствующая точка M, то все точки M заполнят некоторую поверхность (S). Эта поверхность и будет служить «графиком» рассматриваемой функции.

Мы будем пользоваться этим методом, чтобы мысленно представить себе характер изменения рассматриваемой функции; однако значение этого метода ограничено трудностями в практическом исполнении поверхности в пространстве.

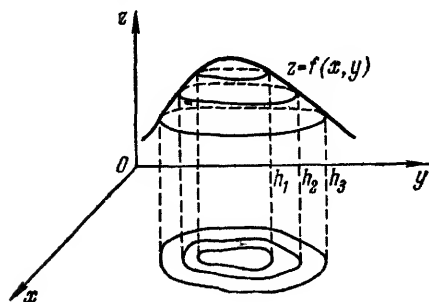


Рис. 191.

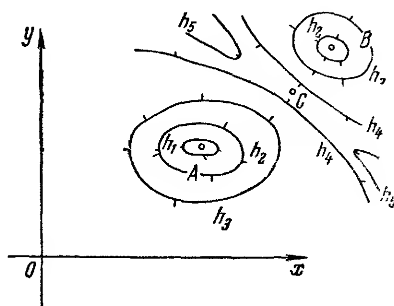


Рис. 192.

Теперь рассмотрим способ, который применяется на практике. Придавая z постоянные значения h_1, h_2, h_3, ..., мы получим в плоскости аргументов линии $f(x, y) = h_1$, $f(x, y) = h_2$, ..., так называемые *линии уровня* функции f. Геометрически они получаются (рис. 191), если пересекать поверхность $z = f(x, y)$ плоскостями, параллельными плоскости xOy, и проектировать линии пересечения на эту плоскость. Этот способ, в частности, широко применяется при черчении географических карт; там функцией служит высота над уровнем океана. Полученная система линий уровня может иметь вид, например, изображенный на рис. 192; маленькие черточки указывают направление убывания функции от линии уровня, для географической карты это — направление стока воды. Из рис. 192 видно, что рассматриваемый график имеет «вершины» в точках A и B (причем в A более высокую), в точке C — *перевал* и т. п.

Имеется специальный раздел математики — *номогра́фия* (от греческих слов «номо» — закон и «графо» — пишу), который изучает методы построения *номогра́мм*, т. е. особых чертежей, служащих для изображения (притом удобного для практического использования) функций от любого числа переменных. Применение номограмм часто влечет значительную экономию времени и средств на вычисления, причем не требует особой квалификации, и поэтому должно быть всячески рекомендовано. Имеется целый ряд типов номограмм. Например, на рис. 193 показана *«номограмма из выравненных точек»* для вычисления одного из углов α_y установки резца на заточном станке по заданным значениям углов резца α и φ. Значения этих величин откладываются

на трех осях, причем если приложить линейку к точкам на осях α и φ, отвечающим заданным значениям, то на третьей оси мы прочитаем искомое значение α_y; например, на рис. 193 показано, что при $\alpha=10°$, $\varphi=30°$ получается $\alpha_y=19{,}5°$.

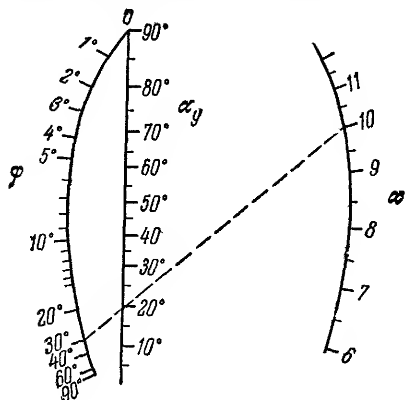


Рис. 193.

2. Область определения. *Область определения функции* $z=f(x, y)$ — это область изменения независимых переменных x, y. Если независимые переменные непрерывные, то ею служит или вся плоскость аргументов, или некоторая ее область, или, наконец, совокупность некоторого числа областей на плоскости x, y. При этом под *областью* на плоскости x, y понимается совокупность точек, *связная*, т. е. состоящая из одного куска, и *не вырождающаяся*, т. е. не линия и, тем более, не точка. Иногда различают области *замкнутые*, если граничная линия причисляется к области, и *открытые* — в противном случае; можно сказать, что области на плоскости играют ту же роль, как интервалы на прямой (п. I.5).

Например, для функции $z=x+y$ область определения — это вся плоскость x, y. Для функции $z=\sqrt{y-x}$, если допускать только вещественные

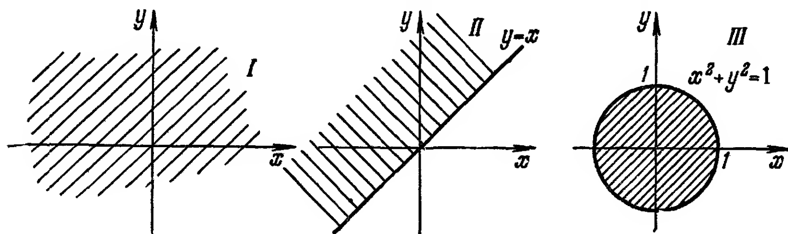


Рис. 194.

значения z, область определения получается из неравенства $y-x\geqslant 0$, т. е. $y\geqslant x$. Для функции $z=\frac{1}{\sqrt{1-x^2-y^2}}$ область определения получается из неравенства $x^2+y^2<1$ и т. п.

Эти области показаны на рис. 194.

3. Линейная функция. *Линейная функция двух переменных* согласно п. I.17 имеет вид

$$z=ax+by+c, \tag{1}$$

где a, b, c — постоянные коэффициенты. Подобно п. I.22, легко получить формулу для ее приращения

$$\Delta z = a\Delta x + b\Delta y;$$

аналогичная формула справедлива для линейной функции любого числа переменных.

Так как формула (1) имеет три коэффициента, то при *линейной аппроксимации*, т. е. при приближенной замене некоторой функции на линейную, требуется три условия. Пусть, например, известны значения некоторой функции $f(x, y)$:

$$f(x_1, y_1) = z_1; \quad f(x_2, y_2) = z_2; \quad f(x_3, y_3) = z_3.$$

Если мы хотим построить линейную функцию (1), принимающую такие же значения, т. е. произвести *линейную интерполяцию*, заменив функцию f на (1), то должно быть

$$\left.\begin{aligned} ax_1 + by_1 + c = z_1; \\ ax_2 + by_2 + c = z_2; \\ ax_3 + by_3 + c = z_3. \end{aligned}\right\} \qquad (2)$$

Из этой системы уравнений мы находим коэффициенты. Такая замена f на (1) дает неплохие результаты внутри треугольника с вершинами $(x_1; y_1)$, $(x_2; y_2)$ и $(x_3; y_3)$ (рис. 195), если он невелик, чтобы не слишком проявлялась нелинейность функции f, и не имеет слишком острых углов. (В пределе, когда один из углов равен нулю, треугольник вырождается в отрезок, а определитель системы (2) обращается в нуль и вычисления неприменимы.)

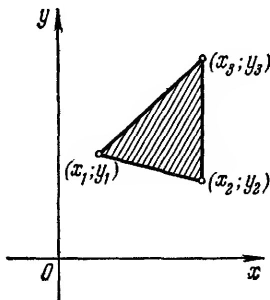


Рис. 195.

При замене f на (1) вне указанного треугольника (*линейная экстраполяция*) ошибка при удалении от него, вообще говоря, возрастает.

Аналогично производится линейная интерполяция в случае любого числа переменных.

4. Непрерывность и разрывы. Понятие непрерывности функции $z = f(x, y)$ совершенно аналогично тому, которое было дано в пп. I.16 и III.12 для случая функции одного переменного. Приведем, например, такое определение непрерывности: функция f называется *непрерывной* при значениях аргументов $x = x_0$, $y = y_0$, если в процессе, когда $x \to x_0$, $y \to y_0$ (произвольным образом), будет $f(x, y) \to f(x_0, y_0)$. В противном случае функция f называется *разрывной* при указанных значениях аргумента, а точка с координатами $(x_0; y_0)$ на плоскости аргументов называется *точкой разрыва* этой функции. Функция, непрерывная в каждой точке некоторой области, называется *непрерывной в этой области*.

Отметим, что наряду с отдельными точками разрыва функции могут иметь целые *линии разрыва*, т. е. линии, целиком состоящие из точек разрыва. Например, из двух функций

$$z = \frac{1}{x^2 + y^2} \quad \text{и} \quad z = \frac{1}{(y - x)^2}$$

первая имеет единственную точку разрыва (0; 0), а вторая — целую линию разрыва (прямую) $(y-x)^2=0$, т. е. $y=x$. Линии уровня этих функций показаны на рис. 196. В обоих случаях в самих точках разрыва функции обращаются в бесконечность. Однако, как и в случае функций одного переменного, бывают и другие виды разрывов. В практических задачах довольно часто встречаются такие линии

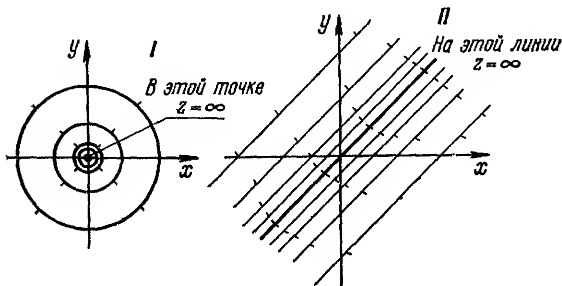


Рис. 196.

разрыва, что при приближении к любой точке этой линии с одной стороны функция имеет один конечный предел, а при приближении к той же точке с другой стороны — другой конечный предел. В этом случае функция имеет при переходе через эту линию конечный скачок; примерный график такой функции изображен на рис. 197.

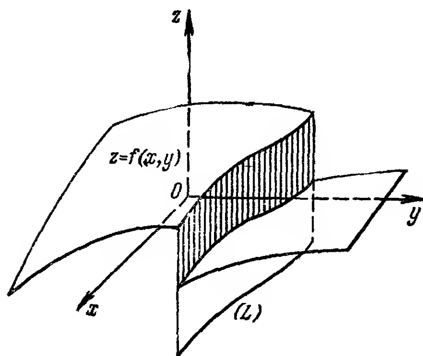


Рис. 197.
(L) — линия разрыва функции f — лежит в плоскости xOy

Поведение функции при приближении к точке разрыва может существенно зависеть от способа этого приближения: так, при приближении по одним путям может существовать предел функции, зависящий от выбора пути, при приближении по другим путям может не существовать ни конечного, ни бесконечного предела и т. д. Так как способов приближения к точке разрыва имеется бесконечное число (тогда как для функций одного переменного имеется лишь два основных способа приближения к точке разрыва — справа или слева), то в целом точки разрыва функций нескольких переменных имеют более сложный вид, чем для функций одной переменной. Например, функция $z=\frac{2xy}{x^2+y^2}$ имеет единственную точку разрыва там, где знаменатель обращается в нуль, т. е. при $x=0$, $y=0$. Если теперь $x\longrightarrow 0$, $y\longrightarrow 0$ и обозначить $y/x=k$, т. е. $y=kx$, то получим $z=\frac{2xkx}{x^2+k^2x^2}=\frac{2k}{1+k^2}$, т. е. предел за-

висит от соотношения между y и x (рис. 198). Если при приближении к некоторой точке нет единого конечного или бесконечного предела функции, то при вычислении предела нужно указывать способ такого приближения.

Свойства непрерывных функций двух переменных на замкнутой конечной области аналогичны тем, которые были описаны в п. III.14 для функций одного переменного на замкнутом конечном интервале, и поэтому мы на них останавливаться не будем.

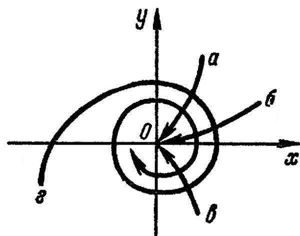


Рис. 198.

При приближении по направлению a предел z равен 1, по направлению $б$ — равен 0, по направлению $в$ — равен -1, по спирали $г$ предела не существует.

Иногда возникает задача о решении неравенств вида $f(x, y) > 0$. Это делается аналогично решению неравенства (III.17); на плоскость x, y надо нанести линию $f(x, y)=0$ и линии разрыва функции f, если они имеются. Все эти линии разобьют плоскость на части, в каждой из которых функция f сохраняет знак; каков этот знак, можно узнать, определяя знак функции в какой-нибудь точке каждой из этих частей.

Решим, например, неравенство

$$\frac{x^2+y^2-4}{x+y} > 0. \qquad (3)$$

Здесь *линией нулей* функции служит окружность $x^2+y^2-4=0$, а линией разрыва — прямая $x+y=0$. Вместе они делят плоскость на четыре части (рис. 199). Беря в каждой из этих частей по точке, например $(-3, 0)$, $(-1; 0)$, $(1; 0)$ и $(3; 0)$, определяем знак функции ($-$, $+$, $-$ и $+$). На рис. 199 области, в которых неравенство (3) справедливо, заштрихованы.

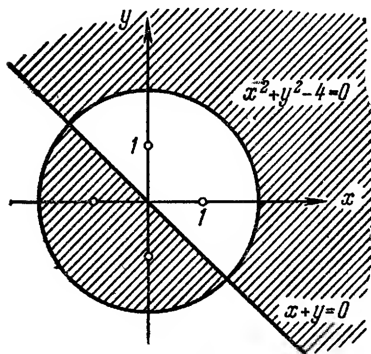


Рис. 199.

5. Неявные функции. Определение неявной функции двух переменных аналогично тому, какое было дано в п. I.20 для функций одной переменной. Неявная функция $z(x, y)$ задается уравнением

$$F(x, y, z) = 0. \qquad (4)$$

Здесь, как и в п. I.20, функция $z(x, y)$ может получиться многозначной, и тогда приходится рассматривать ее однозначные ветви.

Уравнение (4) может определять поверхность любой формы, тогда как поверхность, заданная уравнением $z=f(x, y)$, протыкается любой прямой, параллельной оси z, не более чем в одной точке (см. рис. 190).

§ 2. Функции любого числа переменных

6. Способы задания. Основные понятия, относящиеся к аналитическому виду функции, к ее свойствам, переносятся и на функции любого числа переменных. В то же время при исследовании таких функций имеются дополнительные трудности. Прежде всего табличный и графический способы задания этих функций становятся чрезвычайно громоздкими. Правда, функцию трех переменных можно задавать при помощи системы таблиц с двумя входами (см. п. 1) или набора картинок вида рис. 189 или 192, но и это очень трудно.

Однако в отдельных случаях вычисление значений функции большого числа переменных может быть сведено к вычислению значений нескольких функций меньшего числа переменных. Тогда можно широко применять способы, описанные в пп. *I.13* и *1*. Например, функция четырех переменных вида $u=f(x, y)+\varphi(z, t)$ требует для своего вычисления или наглядного представления таблиц или графиков для функций f и φ, каждая из которых зависит лишь от двух переменных. Аналогично функция четырех переменных вида $u=f[\varphi(x)+y, \psi(z)-t]$ требует задания одной функции двух переменных и двух функций одной переменной и т. п. В этих случаях вычисление и исследование функции значительно облегчаются. К сожалению, далеко не всякую функцию можно представить в подобном виде.

7. Функции трех переменных. Другая трудность возникает при геометрическом истолковании функции, так как в пространстве «не хватает размерности». Сравнительно лучше дело обстоит для функций трех переменных $u=f(x, y, z)$. В этом случае областью определения служит или все *пространство аргументов* x, y, z, или некоторая его часть, т. е. одна или несколько *областей* в пространстве x, y, z (см. п. 2, где понятие невырожденности должно быть естественно изменено), и потому эту область определения можно представить вполне наглядно. Например, для функции $u=x^2+y^2-z$ областью определения служит все пространство, тогда как функция $u=\sqrt{1-x^2-y^2-z^2}$ определена, только если

$$1-x^2-y^2-z^2 \geqslant 0 \quad \text{или} \quad x^2+y^2+z^2 \leqslant 1, \quad |\mathbf{r}| \leqslant 1,$$

т. е. в данном случае областью определения служит шар радиуса 1 с центром в начале координат.

Аналогично п. 1 можно рассматривать *поверхности уровня* функции $f(x, y, z)$, т. е. такие поверхности в пространстве x, y, z, на которых эта функция постоянна, $f(x, y, z)=\text{const}$.

Точки разрыва, если они имеются, находятся в пространстве аргументов и потому их также можно представить наглядно. При этом для функции трех переменных могут быть отдельные точки, линии и даже целые *поверхности разрыва*, т. е. поверхности, целиком состоящие из точек разрыва. Например, при изучении различных сред многие из рассматриваемых величин имеют разрывы на поверхностях раздела (вода — воздух, стекло — воздух и т. п.).

8. Общий случай. Понятие о пространстве аргументов очень наглядное, поэтому желательно сохранить представление об этом пространстве и для случая функций любого числа независимых переменных, даже большего трех. Это делается с помощью многомерного числового пространства (п. VII.18, пример 3). Пусть, например, рассматривается функция четырех переменных $w=f(x, y, z, u)$. Тогда каждый набор значений x, y, z, u определяет точку в пространстве E_4, строго говоря, он и *является* такой точкой. Таким образом, *пространством аргументов* в данном случае служит E_4; областью определения функции f служит область пространства E_4 или совокупность некоторого числа таких областей. Эта функция может быть непрерывной либо иметь отдельные точки, линии, двумерные поверхности или трехмерные гиперповерхности (п. VII.19) разрыва.

«График» функции $u=f(x, y, z, t)$ уже требует пятимерного пространства аргументов и функции x, y, z, t, u; для нахождения точек этого графика надо придавать x, y, z, t произвольные значения и находить соответствующие значения u. [Например, легко проверить, что график функции $u=xz-2y^2t$ проходит через точки (1; 1; 2; 0; 2), (—1; 2; 0, —2; 16) и т. п.] Поэтому в п 7, когда мы говорили о функциях трех переменных, наглядное геометрическое истолкование допускало только пространство аргументов, но не график: он содержится в четырехмерном пространстве E_4, которому мы непосредственно наглядного истолкования не даем.

В п. X. 2 мы увидим, что многомерное пространство допускает менее формальное истолкование, чем числовое пространство.

9. Поле. Говорят, что в *пространстве задано поле некоторой величины, если в каждой точке пространства* или некоторой его области *определено значение этой величины*. Например, при рассмотрении потока газа приходится исследовать температурное поле (в каждой точке температура имеет определенное значение), поле плотностей, поле скоростей и т. д. Поле может быть *скалярным* или *векторным* в зависимости от характера исследуемой величины: например, поля температур или плотностей являются скалярными, а поля скоростей или сил — векторными. Поле может быть *стационарным* (*установившимся*), если в каждой точке пространства оно не меняется с течением времени, или *нестационарным* (*неустановившимся*), если такое изменение имеет место.

Обозначим рассматриваемую, для определенности, скалярную величину буквой u, а произвольную (текущую) точку пространства буквой M. Тогда каждому положению точки M отвечает численное значение величины u, т. е. u можно считать *функцией точки* M, $u=f(M)$. Конечно, это — функция иного рода, чем те, которые мы рассматривали ранее, так как точка не является величиной. Но оказывается, что о функции, в расширенном понимании этого термина, можно говорить всегда, когда имеется определенный закон, согласно которому объектам одного «сорта» (в данном случае точкам пространства) отвечают, соответствуют объекты другого «сорта» (в данном случае значения величины u). Если поле нестационарное, то $u=f(M, t)$.

Легко перейти от функции точки к функции трех переменных — трех пространственных координат. Для этого достаточно ввести декартову систему координат x, y, z в пространство. Тогда положение точки M

в пространстве будет полностью характеризоваться значениями x, y, z, т. е. можно написать $u=u(x, y, z)$; наоборот, в заданной системе координат x, y, z любую функцию от x, y, z можно рассматривать как функцию точки. Однако надо иметь в виду, что поле $u=f(M)$ (функция точки M) имеет смысл и может быть исследовано без всяких систем координат. Кроме того, если вводить системы координат различным способом, то зависимости $u(x, y, z)$ получатся, вообще говоря, различными для одной и той же функции точки $u=u(M)$, т. е. при рассмотрении поля функция точки является *первичной* по отношению к функции координат этой точки.

Если исследуемая величина по своему смыслу задана в плоскости, то соответствующее поле называется *плоским*; такие поля получаются, например, при исследовании тепловых процессов в пластинке, толщиной которой мы пренебрегаем.

Если поле величины u пространственное, но в некоторой декартовой системе координат x, y, z эта величина окажется не зависящей от z, то поле называется *плоскопараллельным*. В таких случаях часто можно отвлечься от координаты z и рассматривать поле только в плоскости x, y, т. е. рассматривать плоскопараллельное поле как плоское, помня о том, что во всех параллельных плоскостях поле имеет в точности тот же вид.

§ 3. Частные производные и дифференциалы первого порядка

10. Основные определения. Пусть дана функция нескольких, например трех, переменных $u=f(x, y, z)$. Если зафиксировать значения всех независимых переменных, кроме одной, то u станет функцией этой одной переменной и по ней можно брать производную или дифференциал, как в гл. IV. Такие *производные* и *дифференциалы* называются *частными*. Другими словами,

$$u'_x=f'_x(x, y, z)=\lim_{\Delta x\to 0}\frac{\Delta_x u}{\Delta x}, \quad u'_y=f'_y(x, y, z)=\lim_{\Delta y\to 0}\frac{\Delta_y u}{\Delta y},$$
$$u'_z=(\text{напишите, чему?}),$$

где $\Delta_x u=f(x+\Delta x, y, z)-f(x, y, z)$, $\Delta_y u=f(x, y+\Delta y, z)-$ $-f(x, y, z)$, $\Delta_z u=$ (чему?) — *частные приращения* функции, полученные за счет изменения одного лишь переменного при фиксированных остальных. Подчеркнем, что *символы u' или $f'(x, y, z)$ для функций нескольких переменных не имеют смысла*, так как надо обязательно указывать, по какой именно переменной производится дифференцирование.

Вычисление частных производных от конкретных элементарных функций осуществляется по правилам п. IV.5, причем в процессе дифференцирования по какой-либо переменной к остальным переменным надо относиться как к постоянным.

Пример. Пусть $u=x^2z^3-y^z$, тогда $u'_x=2xz^3$, $u'_y=-zy^{z-1}$, $u'_z=3x^2z^2-y^z\ln y$ (проверьте!).

Частный дифференциал обозначается символом ∂ со значком (индексом), указывающим, по какой переменной происходит дифференцирование. Таким образом,

$$\partial_x u=u'_x\Delta x;\ \partial_y u=u'_y\Delta y;\ \partial_z u=u'_z\Delta z.$$

Если здесь, в частности, положить $u=x$, то получим

$$\partial_x x=\Delta x,\quad \partial_y x=\partial_z x=0\qquad (\text{так как } x'_x=1,\ x'_y=x'_z=0).\qquad (5)$$

Таким образом, $\partial_x u=u'_x\partial_x x$, откуда $u'_x=\partial_x u/\partial_x x$. В этой формуле обычно не пишут индексов в правой части, т. е. пишут $u'_x=\frac{\partial u}{\partial x}$, поскольку из знаменателя видно, по какому переменному берется дифференциал. Аналогичным образом $u'_y=\frac{\partial u}{\partial y}$, $u'_z=\frac{\partial u}{\partial z}$. Однако в этих формулах различаются не только знаменатели, но и числители: в первой числитель равен $\partial_y u$, а во второй $\partial_z u$. Это правило записи дифференциалов, о котором надо помнить, иногда не очень удобно. Например, вместо $\left(\frac{\partial u}{\partial x}\right)^{-1}$ нельзя написать $\frac{\partial x}{\partial u}$, так как в этих выражениях дифференциалы имеют различный смысл: в первом они берутся по x, а во втором — по u.

При применении частных производных надо тщательно следить за выбором независимых переменных. Например, если записать выражение для мощности электрического тока в виде $P=\frac{U^2}{R}$, где U — напряжение, а R — сопротивление цепи, мы получаем $\frac{\partial P}{\partial R}=-\frac{U^2}{R^2}=-\frac{P}{R}$; если же формулу для мощности записать в виде $P=I^2R$, где I — сила тока, то получаем $\frac{\partial P}{\partial R}=I^2=\frac{P}{R}$. Эти результаты вовсе не противоречат друг другу. В самом деле, первый можно записать подробно в виде $\left.\frac{\partial P}{\partial R}\right|_{U=\text{const}}=-\frac{P}{R}$, а второй — в виде $\left.\frac{\partial P}{\partial R}\right|_{I=\text{const}}=\frac{P}{R}$. (Продумайте физический смысл получающихся знаков.)

11. Полный дифференциал. *Полный дифференциал du функции* $u=f(x, y, z)$ по определению *равен сумме всех ее частных дифференциалов:*

$$du=\partial_x u+\partial_y u+\partial_z u=u'_x\Delta x+u'_y\Delta y+u'_z\Delta z=\frac{\partial u}{\partial x}\Delta x+\frac{\partial u}{\partial y}\Delta y+\frac{\partial u}{\partial z}\Delta z.\qquad (6)$$

Если, в частности, положить $u=x$, то получим в силу формулы (5)

$$dx=\partial_x x+\partial_y x+\partial_z x=\Delta x,$$

т. е. *полный дифференциал независимого переменного равен его приращению* (ср. п. IV.8). Поэтому формулу (6) можно переписать так:

$$du = u'_x dx + u'_y dy + u'_z dz = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy + \frac{\partial u}{\partial z} dz. \qquad (7)$$

Например,

$$d\left(x^2 \sin \frac{x}{y}\right) = \left(2x \sin \frac{x}{y} + \frac{x^2}{y} \cos \frac{x}{y}\right) dx - \frac{x^3}{y^2} \cos \frac{x}{y}\, dy.$$

Связь полного дифференциала с полным приращением функции аналогична той, которая описана в п. IV. 8. Пусть независимые переменные получили приращения Δx, Δy и Δz, в результате чего u получит приращение

$$\Delta u = f(x + \Delta x,\ y + \Delta y,\ z + \Delta z) - f(x,\ y,\ z).$$

Это *полное приращение* можно представить в виде суммы трех приращений, каждое из которых получено за счет изменения только одной переменной (проверьте!):

$$\Delta u = [f(x + \Delta x,\ y,\ z) - f(x,\ y,\ z)] + [f(x + \Delta x,\ y + \Delta y,\ z) - \\ - f(x + \Delta x,\ y,\ z)] + [f(x + \Delta x,\ y + \Delta y,\ z + \Delta z) - \\ - f(x + \Delta x,\ y + \Delta y,\ z)]. \qquad (8)$$

Приращения, стоящие в квадратных скобках, представляют собой частные приращения и поэтому связаны с частными дифференциалами так, как описано в формуле (IV.23):

$$f(x + \Delta x,\ y,\ z) - f(x,\ y,\ z) = f'_x(x,\ y,\ z)\,\Delta x + \alpha_1 \Delta x;$$

$$f(x + \Delta x,\ y + \Delta y,\ z) - f(x + \Delta x,\ y,\ z) = f'_y(x + \Delta x,\ y,\ z)\,\Delta y + \\ + \alpha_2 \Delta y = [f'_y(x,\ y,\ z) + \alpha_3]\,\Delta y + \alpha_2 \Delta y = f'_y(x,\ y,\ z)\,\Delta y + \alpha_4 \Delta y;$$

$$f(x + \Delta x,\ y + \Delta y,\ z + \Delta z) - f(x + \Delta x,\ y + \Delta y,\ z) = (\text{аналогично}) = \\ = f'_z(x,\ y,\ z)\,\Delta z + \alpha_5 \Delta z,$$

где все величины α стремятся к нулю, когда $\Delta x \to 0$, $\Delta y \to 0$ и $\Delta z \to 0$. Подставляя в (8), получим

$$\Delta u = u'_x \Delta x + u'_y \Delta y + u'_z \Delta z + \alpha \Delta x + \beta \Delta y + \gamma \Delta z = \\ = du + \alpha \Delta x + \beta \Delta y + \gamma \Delta z, \qquad (9)$$

где α, β, $\gamma \to 0$ при Δx, Δy, $\Delta z \to 0$. Таким образом, и здесь можно сказать, что *полный дифференциал представляет собой главную линейную часть приращения функции*: главную, так как он отличается от приращения на величины высшего порядка малости (по сравнению с приращениями аргументов), а линейную, так как он представляет собой сумму слагаемых, пропорциональных приращениям аргументов (ср. п. 3). Замена приращения функции на ее дифференциал, как и для функций одного переменного, означает замену нелинейной функции на линейную.

Формула (9) для случая двух независимых переменных проиллюстрирована на рис. 200, где рядом с точками плоскости аргументов указаны значения функции с отброшенными членами высшего порядка малости.

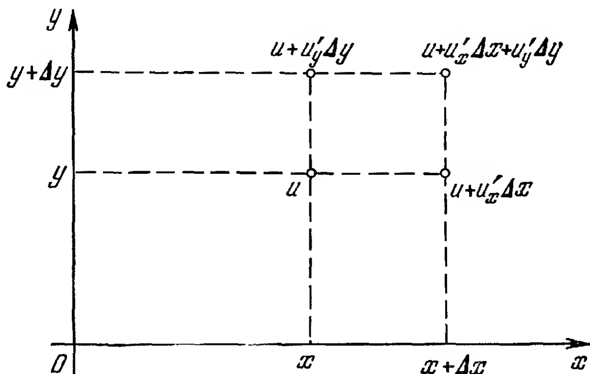


Рис. 200.

Замены $\Delta_x u \approx \partial_x u$, $\Delta_y u \approx \partial_y u$, $\Delta_z u \approx \partial_z u$ и особенно $\Delta u \approx du$, как и в п. IV.10, порождают многочисленные конкретные приближенные формулы. Отметим, что последнюю формулу более подробно можно переписать так:

$$f(a+h,\ b+k,\ c+l) \approx f(a,\ b,\ c) + f'_x(a,\ b,\ c)\,h + f'_y(a,\ b,\ c)\,k + f'_z(a,\ b,\ c)\,l.$$

Аналогично формуле (IV.29) легко получить также, что

$$\alpha_u = |f'_x(\bar{x},\ \bar{y},\ \bar{z})|\,\alpha_x + |f'_y(\bar{x},\ \bar{y},\ \bar{z})|\,\alpha_y + |f'_z(\bar{x},\ \bar{y},\ \bar{z})|\,\alpha_z.$$

Пусть, например, $u = xy$. Тогда

$$\alpha_u = \bar{y}\alpha_x + \bar{x}\alpha_y; \quad \delta_u = \frac{\alpha_u}{\bar{u}} = \frac{\bar{y}\alpha_x + \bar{x}\alpha_y}{\bar{x}\bar{y}} = \frac{\alpha_x}{\bar{x}} + \frac{\alpha_y}{\bar{y}} = \delta_x + \delta_y,$$

т. е. при умножении (и аналогично при делении, проверьте!) приближенных чисел их предельные относительные погрешности складываются.

12. Производная сложной функции. Пусть опять $u = f(x, y, z)$, но переменные x, y, z уже не являются независимыми, а зависят в свою очередь от независимых переменных s, t, т. е.

$$u = u(x,\ y,\ z); \quad x = x(s,\ t); \quad y = y(s,\ t); \quad z = z(s,\ t). \tag{10}$$

Тем самым и u становится сложной функцией s и t. Для вычисления частной производной u'_s зафиксируем t и дадим s приращение Δs. Тогда и x, y, z получат частные приращения, а потому u также получит приращение, которое можно записать по формуле (9) так:

$$\Delta_s u = u'_x \Delta_s x + u'_y \Delta_s y + u'_z \Delta_s z + \alpha \Delta_s x + \beta \Delta_s y + \gamma \Delta_s z.$$

Если теперь разделить обе части на Δs и перейти к пределу

при $\Delta s \to 0$, получим

$$u'_s = \frac{\partial u}{\partial s} = u'_x x'_s + u'_y y'_s + u'_z z'_s = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s}. \qquad (11)$$

Аналогично выражается u'_t. Итак, получающееся правило подобно правилу для функций одного переменного [формула (IV.9)], но число слагаемых здесь больше, так как производные берутся по *всем* промежуточным переменным.

Из формулы (11) аналогично п. IV.9 находим, что формула (7) для полного дифференциала [но не (6)!] остается справедливой и в случае, когда бывшие независимые переменные оказались зависящими от каких-либо других переменных. Действительно, в случае формул (10)

$$du = u'_s ds + u'_t dt = (u'_x x'_s + u'_y y'_s + u'_z z'_s)\, ds + (u'_x x'_t + u'_y y'_t + u'_z z'_t)\, dt =$$
$$= u'_x (x'_s ds + x'_t dt) + u'_y (y'_s ds + y'_t dt) + u'_z (z'_s ds + z'_t dt) =$$
$$= u'_x dx + u'_y dy + u'_z dz,$$

что и требуется. Таким образом, *формула* (7) [равно как и формула (IV.22)] *инвариантна, справедлива во всех случаях.*

Из этой инвариантности получаем многочисленные формулы для дифференцирования. Пусть, например, $w = uv$, где u и v могут зависеть от дальнейших переменных. Тогда по формуле (7) $dw = w'_u du + w'_v dv = v du + u dv$. Итак, формула $d(uv) = v du + u dv$ справедлива во всех случаях. Аналогично проверяется справедливость во всех случаях формул:

$$d(u \pm v) = du \pm dv; \qquad d(Cu) = C du \qquad (C = \text{const});$$
$$d\left(\frac{u}{v}\right) = \frac{v du - u dv}{v^2}; \quad d(u^n) = n u^{n-1} du; \quad d(\sin u) = \cos u\, du \text{ и т. д.}$$

Эти формулы дают возможность во многих случаях вычислять полный дифференциал непосредственно, без вычисления частных производных.

Например, $d(\sin x^2y^3) = \cos x^2y^3\, d(x^2y^3) = \cos(x^2y^3)\,[y^3 d(x^2) + x^2 d(y^3)] = \cos x^2y^3\,(2xy^3dx + 3x^2y^2dy)$; наоборот, по полному дифференциалу можно восстановить частные производные как коэффициенты при dx и dy.

Приведем еще несколько примеров вычисления производных.

1) Пусть $u = f(\sqrt{x^2+y^2})$. Тогда

$$u'_x = f'(\sqrt{x^2+y^2})(\sqrt{x^2+y^2})'_x =$$
$$= f'(\sqrt{x^2+y^2})\,\frac{2x}{2\sqrt{x^2+y^2}} = \frac{x}{\sqrt{x^2+y^2}}\, f'(\sqrt{x^2+y^2}).$$

Здесь функция f сама по себе — функция одной переменной, вместо которой подставлен $\sqrt{x^2+y^2}$; f' означает производную от f по этой единственной переменной.

2) Пусть $u=f\left(\frac{x}{y}, \frac{y}{x}, y\right)$. Тогда

$$u'_y = f'_I\left(\frac{x}{y}, \frac{y}{x}, y\right)\left(-\frac{x}{y^2}\right) + f'_{II}\left(\frac{x}{y}, \frac{y}{x}, y\right)\frac{1}{x} + f'_{III}\left(\frac{x}{y}, \frac{y}{x}, y\right).$$

Здесь функция f сама по себе — функция трех переменных, вместо которых подставлены $\frac{x}{y}$, $\frac{y}{x}$ и y; f'_I, f'_{II} и f'_{III} означают производные от f по этим трем переменным.

3) Пусть $y=x^{\sin x}$. Тогда для вычисления y' в конце п. IV.5 было рекомендовано предварительное логарифмирование. Однако можно вычислять так. Обозначим (обычно это делается в уме) $y=u^{\sin v}$, где $u=x$, $v=x$ и продифференцируем y как сложную функцию, считая u и v промежуточными переменными:

$$y'_x = y'_u u'_x + y'_v v'_x = \sin v\, u^{\sin v - 1}\cdot 1 + u^{\sin v}\ln u \cos v\cdot 1 = \sin x\, x^{\sin x - 1} + x^{\sin x}\ln x\cos x.$$

Ясно, что последний прием является общим. Если нужно вычислить производную по x от выражения, в которое x входит несколько раз, то нужно производить дифференцирование по каждому x, считая остальные зафиксированными, а затем полученные результаты сложить.

Для дальнейшего приведем определение, годное для функций любого числа переменных, хотя мы здесь будем рассматривать только функции трех переменных. Функция $F(x, y, z)$ называется *однородной функцией* измерения k, если для любого $t>0$

$$F(tx, ty, tz) \equiv t^k F(x, y, z). \tag{12}$$

Например, функция $F(x, y, z)=x^2-3yz$ — однородная функция измерения 2, так как

$$F(tx, ty, tz)=(tx)^2-3(ty)(tz)=t^2(x^2-3yz)=t^2F(x, y, z).$$

Аналогично увидим, что функция $\frac{x\sin(y/z)}{y-z}$ — однородная функция нулевого измерения, функция $1/\sqrt{x-y-z}$ — однородная функция измерения $-1/2$, а, например, функция $x+2y-z+1$ — неоднородная функция.

В общем случае (12) при любых a, b, c имеем $F(ta, tb, tc)=t^k F(a, b, c)$. Дифференцируя по t, получим

$$F'_x(ta, tb, tc)\,a + F'_y(ta, tb, tc)\,b + F'_z(ta, tb, tc)\,c = kt^{k-1}F(a, b, c).$$

Полагая в этом равенстве $t=1$, $a=x$, $b=y$, $c=z$, получим *теорему Эйлера об однородных функциях*

$$xF'_x(x, y, z) + yF'_y(x, y, z) + zF'_z(x, y, z) = kF(x, y, z).$$

13. Производные неявных функций. Пусть неявная функция $z=z(x, y)$ определена из уравнения

$$F(x, y, z)=0. \tag{13}$$

Для вычисления частной производной z'_x надо зафиксировать y и дифференцировать формулу (13), имея в виду, что z зависит от x. По правилу дифференцирования сложной функции получим

$$F'_x x'_x + F'_z z'_x = 0, \quad \text{т. е.} \quad F'_x + F'_z z'_x = 0, \tag{14}$$

откуда

$$z'_x = -\frac{F'_x(x, y, z)}{F'_z(x, y, z)}. \tag{15}$$

Аналогично, $z'_y = -F'_y/F'_z$. Если мы хотим, чтобы эта производная принимала определенное конечное значение, то надо требовать, чтобы

$$F'_z(x, y, z) \neq 0. \tag{16}$$

Это—так называемое *условие существования неявной функции* $z = z(x, y)$, определенной из уравнения (13); геометрический смысл этого условия мы покажем в п. XII.3.

Неявные функции могут быть определены из системы уравнений. Пусть дано m уравнений, *совместных* (т. е. непротиворечивых) и *независимых* (т. е. таких, что ни одно из этих уравнений не является следствием остальных), связывающих n переменных. Тогда если $m < n$ (уравнений меньше, чем переменных), то $n - m$ переменных можно принять за независимые, т. е. значения их задавать, а остальные m переменных выражать из заданных m уравнений в виде функций от этих независимых переменных. (Если число уравнений больше или равно числу неизвестных, то для неизвестных получаются дискретные значения и функций не будет). Рассмотрим, например, случай двух уравнений с пятью переменными:

$$\left.\begin{aligned} f(x, y, z, u, v) &= 0, \\ \varphi(x, y, z, u, v) &= 0. \end{aligned}\right\} \tag{17}$$

Здесь можно три переменные принять за независимые, а две другие—за их функции. Примем для определенности $u = u(x, y, z)$, $v = v(x, y, z)$ и попробуем найти производные u'_x и v'_x, для этого произведем дифференцирование обоих уравнений (17) при фиксированных y и z. Получим

$$\left.\begin{aligned} f'_x \cdot 1 + f'_u u'_x + f'_v v'_x &= 0, \\ \varphi'_x \cdot 1 + \varphi'_u u'_x + \varphi'_v v'_x &= 0, \end{aligned}\right\} \quad \text{т. е.} \quad \left.\begin{aligned} f'_u u'_x + f'_v v'_x &= -f'_x; \\ \varphi'_u u'_x + \varphi'_v v'_x &= -\varphi'_x. \end{aligned}\right\} \tag{18}$$

Таким образом, относительно u'_x и v'_x получилась система двух уравнений первой степени с двумя неизвестными. Для ее разрешимости нужно (см. п. VI.4), чтобы определитель системы был отличен от нуля:

$$\begin{vmatrix} f'_u & f'_v \\ \varphi'_u & \varphi'_v \end{vmatrix} \neq 0. \tag{19}$$

Аналогично находятся производные по y и z; определитель системы будет тот же (изменятся правые части). Таким образом, (19)—это *условие существования неявных функций* $u = u(x, y, z)$ и $v = v(x, y, z)$, определенных из системы уравнений (17). Аналогичный вид имеет условие для любого числа уравнений и переменных.

Функциональный определитель (т. е. определитель, составленный из функций) вида (19) часто встречается в математике и имеет специальное название *«якобиа́н»* (по имени немецкого математика К. Яко́би (1804—1851)) и симво-

лическое обозначение $\frac{D(f,\ \varphi)}{D(u,\ v)}$; такой символ надо понимать как единый, так как числитель и знаменатель в отдельности пока для нас не имеют смысла.

Аналогичные вопросы возникают при решении *систем уравнений, включающих параметры*. Пусть, например, рассматривается система двух уравнений с двумя неизвестными x, y:

$$\left.\begin{array}{l} f(x,\ y;\ \alpha,\ \beta,\ \gamma,\ \ldots)=0, \\ \varphi(x,\ y;\ \alpha,\ \beta,\ \gamma,\ \ldots)=0, \end{array}\right\}$$

где α, β, γ, ...— параметры. Если при некоторых значениях α_0, β_0, γ_0, ... этих параметров система имеет решение x_0, y_0 и если при этих значениях $\frac{D(f,\ \varphi)}{D(x,\ y)} \neq 0$, то в силу сказанного система определяет x и y как функции α, β, γ, т. е. при изменении параметров система будет продолжать иметь вполне определенное решение. Правда, как будет пояснено в п. XII.3, это утверждение имеет локальный характер, т. е. при значительном изменении параметров система может перестать быть разрешимой. Подчеркнем, что для такой «устойчивости» решения относительно изменения параметров число m уравнений в системе должно равняться числу n неизвестных. Если $m < n$, то $n-m$ неизвестных остаются произвольными, а если $m > n$, то для разрешимости требуется выполнение $m-n$ соотношений между параметрами.

§ 4. Частные производные и дифференциалы высших порядков

14. Определения. Пусть для определенности $u=f(x,\ y,\ z)$; аналогично рассматриваются функции любого числа переменных. Тогда, как мы видели, имеются три частные производные первого порядка: $u'_x = f'_x(x,\ y,\ z)$, u'_y и u'_z. Каждую из них можно вновь дифференцировать по x, y и z, так что получается *девять частных производных второго порядка*:

$$u''_{xx}=f''_{xx}(x,\ y,\ z);\quad u''_{xy}=f''_{xy}(x,\ y,\ z);\quad u''_{xz};\quad u''_{yx};\quad u''_{yy};\quad u''_{yz};\quad u''_{zx};\quad u''_{zy}\quad \text{и}\quad u''_{zz}.$$

Дифференцирование элементарных функций, заданных в явном виде, осуществляется по правилам п. IV.5, а неявных функций — при помощи дальнейшего дифференцирования равенств (14), (15) или (18) и т. п. Производные более высокого порядка определяются аналогично.

Подобным образом определяются частные дифференциалы высших порядков и подобно пп. IV.12 и 10 мы приходим к равенствам

$$\partial^2_{xx} u = u''_{xx}\,dx^2;\quad \partial^2_{xy} u = u''_{xy}\,dx\,dy \quad \text{и т. д.,} \tag{20}$$

где дифференциал независимой переменной $dx=\Delta x=\partial_x x$ и т. п. Отсюда

$$u''_{xx}=\frac{\partial^2_{xx} u}{(\partial_x x)^2}=\frac{\partial^2 u}{\partial x^2};\quad u''_{xy}=\frac{\partial^2_{xy} u}{\partial_x x \partial_y y}=\frac{\partial^2 u}{\partial x\,\partial y};\quad u''_{xz}=\frac{\partial^2 u}{\partial x\,\partial z}$$

и т. д.

Подобно п. V.7 можно определить *частные разности* для функций нескольких переменных, однако при этом непременно надо указывать, по какой именно переменной берется разность, причем по разным переменным шаг может быть различным. Пусть, например, $z=f(x, y)$; тогда можно обозначить шаг по x через h и под Δ_h понимать частную разность по x, т. е. $\Delta_h z=f(x+h, y)-f(x, y)$, а шаг по y обозначить через k и под Δ_k понимать частную разность по y, т. е. $\Delta_k z=f(x, y+k)-f(x, y)$. Тогда естественно обозначать

$$\Delta^2_{hh}\, z=\Delta_h(\Delta_h z), \quad \Delta^2_{hk}\, z=\Delta_k(\Delta_h z) \text{ и т. п.}$$

Связь между раздельными разностями и производными выражается формулами

$$z'_x=\lim_{h\to 0}\frac{\Delta_h z}{h}; \quad z'_y=\lim_{k\to 0}\frac{\Delta_k z}{k}; \quad z'_{xx}=\lim_{h\to 0}\frac{\Delta^2_{hh}\, z}{h^2}; \quad z''_{xy}=\lim_{\substack{h\to 0\\ k\to 0}}\frac{\Delta^2_{hk}\, z}{hk}$$

и т. д

15. Равенство смешанных производных. Пусть $z=f(x, y)$. Тогда у этой функции имеется четыре частные производные второго порядка: z''_{xx}, z''_{xy}, z''_{yx}, z''_{yy}. Оказывается, что две средние производные, которые называются *смешанными*, равны между собой:

$$z''_{xy}=z''_{yx} \tag{21}$$

т. е. *смешанные производные не зависят от того, в каком порядке производится дифференцирование.*

Для доказательства достаточно заметить, что согласно концу п. 14

$$z''_{xy}=\lim_{h,k\to 0}\frac{1}{hk}\Delta^2_{hk}\, z, \quad z''_{yx}=\lim_{h,k\to 0}\frac{1}{hk}\Delta^2_{kh}\, z. \tag{22}$$

В то же время

$$\Delta^2_{hk}\, z=\Delta_k(\Delta_h z)=\Delta_k[f(x+h, y)-f(x, y)]=$$
$$=[f(x+h, y+k)-f(x, y+k)]-[f(x+h, y)-f(x, y)]=$$
$$=f(x+h, y+k)-f(x, y+k)-f(x+h, y)+f(x, y);$$
$$\Delta^2_{kh}\, z=\Delta_h(\Delta_k z)=\Delta_h[f(x, y+k)-f(x, y)]=[f(x+h, y+k)-f(x+h, y)]-$$
$$-[f(x, y+k)-f(x, y)]=f(x+h, y+k)-f(x+h, y)-f(x, y+k)+f(x, y),$$

т. е.

$$\Delta^2_{hk}\, z=\Delta^2_{kh}\, z$$

—*смешанные разности не зависят от того, в каком порядке берутся разности.* Отсюда из (22) и вытекает формула (21).

Если теперь рассматриваются производные еще более высокого порядка, то согласно формуле (21) можно менять порядок любых двух рядом стоящих дифференцирований, в результате чего можно от любого порядка дифференцирований перейти к любому другому; существенно только то, сколько раз по какому переменному производится дифференцирование, но не то, в каком порядке. Например,

$$u^{\mathrm{IV}}_{xxyz}=u^{\mathrm{IV}}_{xyxz}=u^{\mathrm{IV}}_{xyzx}=u^{\mathrm{IV}}_{xzyx}=u^{\mathrm{IV}}_{zxyz} \quad \text{и т. д., но} \quad \neq u^{\mathrm{IV}}_{xyyz}.$$

Частные дифференциалы [см. формулу (20)] также не зависят от порядка, в котором производится дифференцирование.

16. Полный дифференциал высшего порядка. Полный дифференциал любого порядка — это полный дифференциал (см. п. 11) от полного дифференциала предыдущего порядка. Как и ранее (п. IV.12), при последующих дифференцированиях дифференциалы независимых переменных надо рассматривать как постоянные величины. Пусть, например, $z=f(x, y)$. Тогда

$$dz = z'_x\,dx + z'_y\,dy;$$

$$\begin{aligned} d^2z = d(dz) &= (z'_x\,dx + z'_y\,dy)'_x\,dx + (z'_x\,dx + z'_y\,dy)'_y\,dy = \\ &= z''_{xx}\,dx^2 + z''_{yx}\,dy\,dx + z''_{xy}\,dx\,dy + z''_{yy}\,dy^2 = \\ &= z''_{xx}\,dx^2 + 2z''_{xy}\,dx\,dy + z''_{yy}\,dy^2; \end{aligned} \tag{23}$$

при этом мы использовали формулу (21). Далее,

$$\begin{aligned} d^3z &= (z''_{xx}\,dx^2 + 2z''_{xy}\,dx\,dy + z''_{yy}\,dy^2)'_x\,dx + (z''_{xx}\,dx^2 + 2z''_{xy}\,dx\,dy + \\ &+ z''_{yy}\,dy^2)'_y\,dy = (z'''_{xxx}\,dx^3 + 2z'''_{xxy}\,dx^2\,dy + z'''_{xyy}\,dx\,dy^2) + \\ &+ (z'''_{xxy}\,dx^2\,dy + 2z'''_{xyy}\,dx\,dy^2 + z'''_{yyy}\,dy^3) = \\ &= z'''_{xxx}\,dx^3 + 3z'''_{xxy}\,dx^2\,dy + 3z'''_{xyy}\,dx\,dy^2 + z'''_{yyy}\,dy^3. \end{aligned}$$

Как и при выводе формулы (IV.32), мы видим, что вычисления идут по той же схеме, как если последовательно раскрыть скобки в выражениях $(a+b)^2$, $(a+b)^3$ и т. д. В общем случае

$$d^nz = \frac{\partial^n z}{\partial x^n}\,dx^n + \binom{n}{1}\frac{\partial^n z}{\partial x^{n-1}\,\partial y}\,dx^{n-1}\,dy + \binom{n}{2}\frac{\partial^n z}{\partial x^{n-2}\,\partial y^2}\,dx^{n-2}\,dy^2 + \ldots + \frac{\partial^n z}{\partial y^n}\,dy^n.$$

Этот результат можно записать в виде символической формулы

$$d^nz = \left(dx\,\frac{\partial}{\partial x} + dy\,\frac{\partial}{\partial y}\right)^n z,$$

где в правой части надо раскрыть скобки так, как если бы ∂, ∂x, ∂y, dx и dy были обычными алгебраическими множителями. Подобным образом, если $u=f(x, y, z)$, то $d^nu = \left(dx\,\frac{\partial}{\partial x} + dy\,\frac{\partial}{\partial y} + dz\,\frac{\partial}{\partial z}\right)^n u$ и т. п.

Если $z=f(x, y)$, но x и y не являются независимыми переменными, то, подобно концу п. IV.12, формулу (23) надо изменить:

$$\begin{aligned} d^2z &= d(z'_x\,dx + z'_y\,dy) = d(z'_x\,dx) + d(z'_y\,dy) = d(z'_x)\,dx + z'_x\,d(dx) + d(z'_y)\,dy + \\ &+ z'_y\,d(dy) = (z''_{xx}\,dx + z''_{xy}\,dy)\,dx + z'_x\,d^2x + (z''_{yx}\,dx + z''_{yy}\,dy)\,dy + z'_y\,d^2y = \\ &= z''_{xx}\,dx^2 + 2z''_{xy}\,dx\,dy + z''_{yy}\,dy^2 + z'_x\,d^2x + z'_y\,d^2y. \end{aligned} \tag{24}$$

Аналогично изменятся выражения для дальнейших дифференциалов.

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ В ПРОСТРАНСТВЕ

§ 1. Координаты в пространстве

1. **Различные виды координат в пространстве.** Помимо декартовых координат, описанных в п. VII.9, широко применяются следующие системы координат.

1. *Цилиндрические координаты* (ρ; φ; z), показанные на рис. 201,— это полярные координаты на плоскости (см. п. II.3), к которым добавлена координата z. Ясно, что для описания всех точек в пространстве достаточно значений $0 \leqslant \rho < \infty$; $-\pi < \varphi \leqslant \pi$; $-\infty < z < \infty$.

Координатные поверхности, т. е. поверхности, на которых одна из координат постоянна, а остальные две меняются, образуют три семейства: $\rho = \mathrm{const}$, $\varphi = \mathrm{const}$ и $z = \mathrm{const}$. На рис. 202 показано по одному представителю из этих семейств; конечно, все поверхности надо считать продолженными в бесконечность. *Координатные линии*, на которых две координаты постоянны, а одна меняется, также образуют три семейства и получаются в результате пересечения координатных поверхностей; они показаны на рис. 202 жирными линиями. (Заметим, что для декартовой системы координат координатными поверхностями служат плоскости, параллельные одной из плоскостей xOy, yOz или zOx, а координатными линиями — прямые, параллельные одной из координатных осей.)

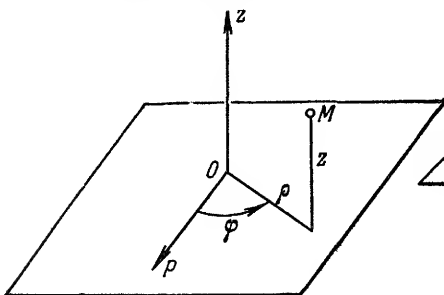


Рис. 201.

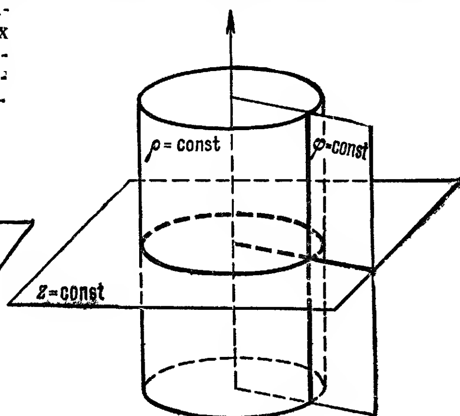


Рис. 202.

Связь между декартовыми координатами (x; y; z) и цилиндрическими (ρ; φ; z), если обе системы координат расположены друг относительно друга так, как показано на рис. 203, выражается формулами $x = \rho\cos\varphi$, $y = \rho\sin\varphi$, $z = z$.

Цилиндрические координаты часто применяются при рассмотрении тел вращения (круговые цилиндр и конус и т. д.), причем ось z располагается по оси вращения.

2. *Сферические координаты* (r; ϑ; φ), показанные на рис. 204, аналогичны географическим с той разницей, что «широта» ϑ здесь отсчитывается не от экватора, как в географии, а от «северного полюса». Интервалы изменения координат, достаточные для описания всех точек в пространстве, таковы: $0 \leqslant r < \infty$; $0 \leqslant \vartheta \leqslant \pi$; $-\pi < \varphi \leqslant \pi$.

Координатные поверхности и линии показаны на рис. 205. Связь между декартовыми координатами (x; y; z) и сферическими (r; ϑ; φ), если системы координат расположены, как на рис. 206, выражается формулами

$$x = \rho \cos \varphi = r \sin \vartheta \cos \varphi;$$
$$y = r \sin \vartheta \sin \varphi; \quad z = r \cos \vartheta.$$

Сферические координаты особенно удобны при рассмотрении тел, ограни-

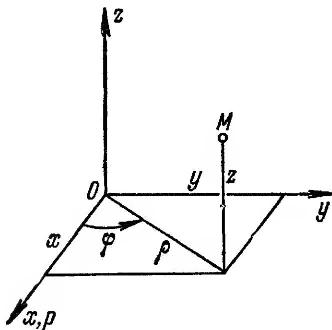


Рис. 203.

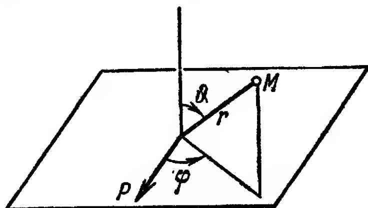


Рис. 204.

ченных поверхностями, показанными на рис. 205, но применяются и во многих других случаях.

Декартовы, цилиндрические и сферические координаты — это частные случаи *ортогональных координат*, характерных тем, что угол между координатными линиями в точках их пересечения прямой (проверьте!). Иногда применяются и неортогональные координаты, например общие аффинные (п. VII.9).

2. **Число степеней свободы.** Мы видели, что в пространстве возможны различные системы координат, но для всех них является общим то, что положение

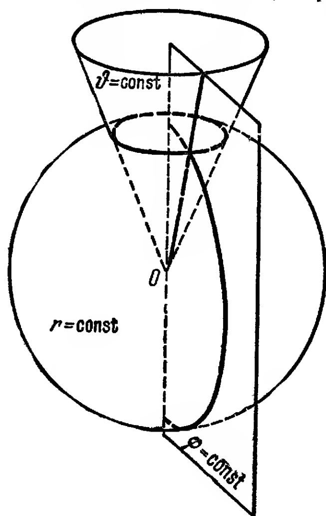


Рис. 205.

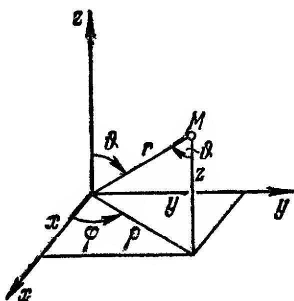


Рис. 206.

точки в пространстве определяется *тремя* координатами, тогда как положение точки на плоскости определяется *двумя* координатами, а на линии — *одной*. Это выражают словами: при выборе точки в (геометрическом) прост-

ранстве или, что то же, при движении точки в пространстве имеются *три степени свободы*, тогда как при выборе точки на плоскости (а также на любой поверхности) имеются две степени свободы, а на линии — одна. Или, другими словами: пространство трехмерно, тогда как поверхности двумерны, а линии одномерны.

В общем случае понятие о числе степеней свободы вводится так. Пусть имеется некоторая совокупность объектов (в предыдущем примере — совокупность всех точек в пространстве), каждый из которых может быть охарактеризован указанием численных значений некоторых параметров (в предыдущем примере — координат). Пусть эти параметры являются:

1) *независимыми*, т. е. могут принимать произвольные значения: например, если зафиксировать все параметры, кроме одного, то этот один можно еще произвольно менять, быть может, в некоторых пределах;

2) *существенными*, т. е. при любом изменении параметров рассматриваемый объект фактически меняется. Тогда, если таких параметров k, говорят, что при выборе объекта из рассматриваемой совокупности имеется *k степеней свободы*, а сама совокупность называется (обобщенным) *k-мерным пространством* или *k-мерным многообразием*. Сами параметры называются (обобщенными) *координатами* в этом пространстве; как и в случае обычных координат в обычном пространстве, их можно выбирать различным способом, как это окажется удобнее в том или ином исследовании. Объекты, составляющие пространство, называются его *элементами* или *точками*. Таким образом, многомерное пространство получает конкретное истолкование.

Приведенное определение размерности согласуется с определением размерности линейного пространства, данным в п. VII. 19, так как за параметры в таком пространстве можно принять коэффициенты разложения вектора по некоторому фиксированному базису. Но теперь мы рассматриваем пространства значительно более общего вида, элементы которых объединены только «общей природой» и понятием близости между элементами, которой должна отвечать близость между соответствующими параметрами. Такие общие пространства с понятием близости между элементами или, что то же, с понятием перехода к пределу называются *топологическими пространствами*.

Приведем несколько примеров. Пусть рассматривается совокупность всевозможных кругов на плоскости. Каждый круг полностью характеризуется тремя параметрами: координатами $(x;\ y)$ центра и радиусом r. Эти параметры независимы (их можно произвольно менять) и существенны (при их изменении круг также меняется). Таким образом, при выборе круга на плоскости имеются три степени свободы, т. е. такая совокупность кругов образует трехмерное пространство с обобщенными координатами $(x;\ y;\ r)$. Аналогично совокупность всевозможных шаров в пространстве образует четырехмерное пространство.

В физике систематически рассматривается совокупность *событий*, каждое из которых полностью характеризуется ответами на вопросы «где?» и «когда?» На первый вопрос можно ответить указанием, например, декартовых координат x, y, z, а на второй — указанием момента времени t. Таким образом, пространство событий — четырехмерное и обобщенными координатами в нем могут служить $(x;\ y;\ z;\ t)$.

Еще пример: сколько степеней свободы имеет отрезок данной длины l при движении в пространстве? Каждый такой отрезок полностью определяется координатами $(x_1;\ y_1;\ z_1)$ и $(x_2;\ y_2;\ z_2)$ его концов; эти координаты можно принять за параметры, определяющие положение отрезка. Эти параметры, очевидно, существенные, однако они не являются независимыми, а связаны соотношением (см. формулу (VII. 14))

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}=l.$$

Таким образом, только пять параметров можно считать независимыми, а

шестой выражается через них из этого соотношения. Отрезок данной длины при движении в пространстве имеет пять степеней свободы.

В общем случае, если параметров n и они существенные, но связаны m *независимыми* уравнениями (т. е. такими уравнениями, из которых ни одно не вытекает из остальных), то $n-m$ параметров можно принять за независимые, а остальные m будут через них выражаться, т. е. будет $n-m$ степеней свободы. Таким образом, например, при движении треугольника в пространстве получится $9-3=6$ (проверьте!) степеней свободы. Этот пример важен в связи с тем, что положение абсолютно твердого тела произвольной формы полностью определяется указанием положения трех его точек, не лежащих на одной прямой (почему?). Значит, при движении такого тела в пространстве также имеются шесть степеней свободы.

Найдем еще число степеней свободы при движении бесконечной прямой в пространстве. Можно рассуждать так: выберем произвольно две точки A и B в пространстве (каждая имеет по три координаты) и проведем через них прямую (P), которая будет определяться, таким образом, шестью параметрами. Так как эти параметры независимые, то, казалось бы, получается шесть степеней свободы. Однако такое рассуждение неверно, так как при изменении этих параметров точки A и B будут, правда, меняться, но прямая (P) может при этом оставаться неизменной; значит, требование существенности параметров не выполняется. Так как прямая (P) не меняется, если точка A скользит по ней (одна степень свободы) или точка B скользит по ней (еще одна степень свободы), то при нашем подсчете получились две лишние степени свободы и на самом деле число степеней свободы равно $6-2=4$. За независимые и существенные параметры можно взять, например, координаты точек пересечения прямой (P) с плоскостями xOy и yOz; правда, не все прямые пересекаются с этими плоскостями, но эти особые случаи не могут сказаться при подсчете числа степеней свободы.

В k-мерном пространстве могут быть совокупности точек, т. е. *подпространства* (многообразия) той же, либо меньшей размерности. Попадание на многообразие (S) размерности $k-1$ является исключительным обстоятельством, если, конечно, это попадание специально не предусмотрено; оно обусловливается выполнением определенного соотношения вида $f_S(\alpha_1, \alpha_2, \ldots, \alpha_k)=0$ между координатами $\alpha_1, \alpha_2, \ldots, \alpha_k$. Типичным, основным случаем является непопадание на (S), т. е. выполнение неравенства $f_S \neq 0$. Если это неравенство выполнено при некоторых значениях координат, то оно сохраняется при любых достаточно близких значениях, тогда как равенство $f_S=0$ может при как угодно малом изменении координат нарушиться. Поэтому говорят, что свойство, характеризуемое неравенствами между координатами, является *грубым, структурно устойчивым* относительно изменения координат, тогда как свойство, характеризуемое равенствами, является *тонким, неустойчивым.* Однако, если координаты меняются так, что f_S непрерывно переходит от отрицательных значений к положительным, то по пути должно быть и $f_S=0$, т. е. точка попадает на (S). Еще более трудным является попадание на многообразие размерности $k-p<k-1$; для этого должно выполняться p соотношений вида равенств.

Рассмотрим, например, систему (VI.1) двух уравнений первой степени с двумя неизвестными. Пространство таких систем шестимерно, так как система определяется параметрами a_1, b_1, d_1, a_2, b_2, d_2, которые и могут считаться координатами системы. Особый случай (п. VI.6) определяется равенством

$$D=a_1b_2-a_2b_1=0,$$

т. е. структурно устойчивым является основной случай $D \neq 0$, а подпространство (S) особых случаев пятимерно. Среди этих особых случаев типичным является несовместность системы, тогда как системы с бесконечным числом решений образуют четырехмерное подпространство в (S) (почему?).

В заключение отметим, что бывают случаи, когда параметры могут принимать произвольные комплексные значения; тогда говорят о *комплексной размерности*. Так как у каждого комплексного параметра произвольны и вещественная, и мнимая части, то комплексная k-мерность означает вещественную $2k$-мерность.

§ 2. Поверхности и линии в пространстве

3. Поверхности в пространстве. В п. IX.5 мы показали, что уравнение

$$F(x, y, z) = 0 \tag{1}$$

определяет в пространстве x, y, z (т. е. в пространстве, в котором задана декартова система координат x, y, z) некоторую поверхность (S), представляющую собой совокупность точек, координаты которых удовлетворяют данному уравнению (1). При этом соотношение (1) называется *уравнением поверхности* (S). Если, наоборот, первоначально дана поверхность (S) в пространстве x, y, z, то можно получить ее уравнение в форме (1). Например, рассуждая аналогично п. II.4 и опираясь на формулу (VII.14), легко вывести уравнение сферы с центром в точке $(a; b; c)$ и радиусом R:

$$(x-a)^2 + (y-b)^2 + (z-c)^2 - R^2 = 0.$$

Уравнение поверхности можно записать и в других координатах: например, в сферических оно имеет вид $\Phi(r, \vartheta, \varphi) = 0$.

Аналогично п. II.4 при отыскании точек пересечения трех поверхностей, уравнения которых заданы в форме (1), приходится решать систему трех уравнений с тремя неизвестными:

$$\left.\begin{aligned} F_1(x, y, z) &= 0; \\ F_2(x, y, z) &= 0; \\ F_3(x, y, z) &= 0. \end{aligned}\right\}$$

Подобно п. II.7 вводится понятие об *алгебраических* и *трансцендентных поверхностях*. Как и в п. II.8, возможны особые случаи: мнимые поверхности, случаи *вырождения* и *распадения*. Следует только иметь в виду, что в п. II.8 линия могла вырождаться в точку, а поверхность может вырождаться в точку (например, «сфера нулевого радиуса») или в линию; например, бесконечный «круговой цилиндр нулевого диаметра» представляет собой прямую линию.

4. Цилиндры, конусы, поверхности вращения. Возьмем для примера уравнение $z - x^2 = 0$. Если его рассматривать только в плоскости xOz, то это уравнение параболы (L) $(z = x^2)$, которой принадлежат точки $O(x=0;\ z=0)$; $A(x=2;\ z=4)$ и т. п. Но если это же уравнение рассматривать во всем пространстве x, y, z, то получится уравнение цилиндрической поверхности (рис. 207), направляющей которой служит парабола (L), а образующие парал-

лельны оси y (*параболический цилиндр*). Действительно, наряду, например, с точкой A рассматриваемой поверхности принадлежат точки (2; 5; 4), (2; —8; 4) и вообще все точки с координатами (2; y; 4) при произвольном y, так как эти координаты удовлетворяют рассматриваемому уравнению (оно не содержит y, т. е. y может быть любым, лишь бы x и z были такими, как нужно). Но эти точки заполняют целую прямую, показанную на рис. 207 жирно; аналогично рассматриваются и остальные точки параболы (L).

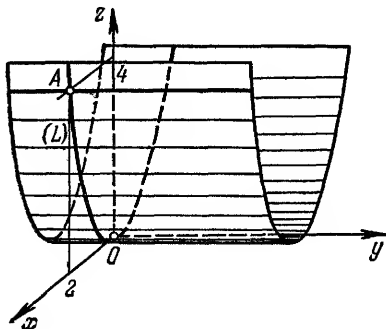


Рис. 207.

Подобным образом любое уравнение вида $F(x, z)=0$ представляет собой в пространстве x, y, z уравнение некоторой *цилиндрической поверхности* с образующими, параллельными оси y, и направляющей в плоскости xOz с тем же уравнением $F(x, z)=0$. Уравнения же $\Phi(x, y)=0$ и $\Psi(y, z)=0$ представляют цилиндрические поверхности с образующими, параллельными соответственно оси z или оси x. Например, уравнение $x^2+y^2=R^2$ представляет прямой круговой цилиндр радиуса R с осью Oz (то же уравнение в плоскости xOy представляет окружность).

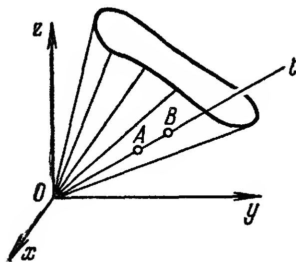


Рис. 208.

Рассмотрим теперь уравнение (1), причем будем предполагать, что функция F однородная (см. конец п. IX.12), и докажем, что тогда это уравнение представляет *коническую поверхность* с вершиной в начале координат. Действительно (рис. 208), пусть рассматриваемая поверхность содержит некоторую точку A с координатами ($\bar{x}$; $\bar{y}$; $\bar{z}$). Тогда $F(\bar{x}; \bar{y}; \bar{z})=0$, так как координаты точки A должны удовлетворять уравнению поверхности. Возьмем теперь любую точку B с координатами ($t\bar{x}$; $t\bar{y}$; $t\bar{z}$), где t — какое угодно положительное число. Тогда

$$F(t\bar{x};\ t\bar{y};\ t\bar{z})=t^kF(\bar{x},\ \bar{y},\ \bar{z})=t^k0=0,$$

т. е. и точка B лежит на рассматриваемой поверхности. Но если теперь менять t, то точка B опишет целый луч l, который, таким образом, тоже лежит на этой поверхности. Итак, рассматриваемая поверхность вместе с любой своей точкой A содержит целый луч l, откуда вытекает, что эта поверхность — коническая, точнее,

представляет собой «*полуконус*»; конус получится, если t в тождестве (IX.12) может быть и отрицательным. Например, уравнение $ax^2 + by^2 - cz^2 = 0$ $(a, b, c > 0)$ представляет конус; так как в плоскости $z = 1$ получается эллипс (проверьте!), то это — эллиптический конус с осью Oz.

В заключение рассмотрим уравнения *поверхностей вращения*. Пусть например, линия (L), лежащая в плоскости yOz и имеющая уравнение $F(y, z) = 0$, вращается относительно оси z (рис. 209); выведем уравнение полученной поверхности. Для этого возьмем на ней произвольную точку $M(x; y; z)$ и найдем соответствующую точку $\overline{M}(\bar{x}; \bar{y}; \bar{z})$ на линии (L). Тогда $\bar{z} = z$, $\bar{x} = 0$, а для вычисления $\bar{y}$ заметим, что $\bar{y} = K\overline{M} = KM = \sqrt{x^2 + y^2}$ (проверьте!). Так как точка $\overline{M}$ лежит на (L), то $F(\bar{y}, \bar{z}) = 0$, т. е. $F(\sqrt{x^2 + y^2}, z) = 0$. Это и есть уравнение рассматриваемой поверхности вращения. Например, $z = ay^2$ — уравнение параболы в плоскости yOz, а $z = a(x^2 + y^2)$ — уравнение параболоида вращения.

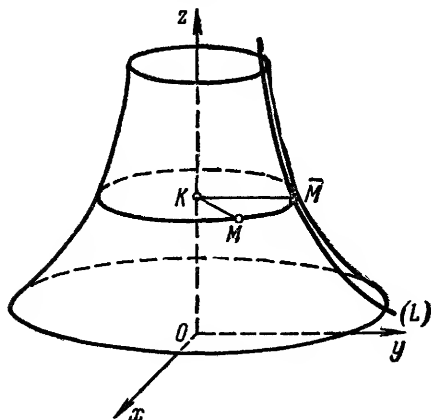


Рис. 209.

5. Линия в пространстве. Линия в пространстве может получиться как результат пересечения (т. е. как общая часть) двух поверхностей или как след (траектория) движущейся точки. В первом случае уравнения обеих поверхностей в декартовых координатах можно записать в виде

$$\left.\begin{aligned} F_1(x, y, z) = 0; \\ F_2(x, y, z) = 0. \end{aligned}\right\} \tag{2}$$

Тогда, поскольку точки линии их пересечения принадлежат одновременно обеим поверхностям, эта линия будет представлять собой совокупность точек, координаты которых удовлетворяют одновременно обоим уравнениям (2), т. е. (2) надо рассматривать как систему двух уравнений с тремя неизвестными.

При втором подходе уравнение линии получается в параметрическом виде

$$x = \varphi(t); \quad y = \psi(t); \quad z = \chi(t) \tag{3}$$

(см. п. VII.23). Для перехода от этого вида к виду (2) надо из уравнений (3) исключить t (например, из первого уравнения выразить t через x и подставить результат в два других уравнения), если это удастся и если это целесообразно. Для обратного перехода

от (2) к (3) надо (при тех же «если») в (2) подставить, например, $x=t$, после чего разрешить эти два уравнения относительно y и z; в результате y и z выразятся через t.

Иногда возникает задача о нахождении проекции (L') заданной линии (L) на одну из координатных плоскостей, например на плоскость xOy (рис. 210). Это означает, что требуется найти соотношение между x и y для точек этой линии. Если (L) задана уравнениями (2), то для нахождения (L') из них надо исключить z, а если (L) задана в виде (3), то надо просто оставить два первых равенства.

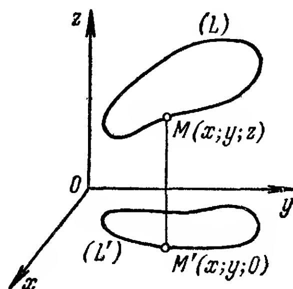


Рис. 210.

6. Параметрическое задание поверхностей в пространстве и функций. В п. 5 мы видели (формулы (3)), что для параметрического представления линии требуется один параметр. Посмотрим теперь, что получится, если параметров два, т. е. каков геометрический смысл уравнений

$$x=\varphi(u, v), \quad y=\psi(u, v), \quad z=\chi(u, v), \qquad (4)$$

в которых параметры u и v принимают произвольные числовые значения. Естественно предположить, что эти уравнения представляют поверхность в пространстве, так как точка на поверхности имеет две степени свободы и потому требует для своего указания двух параметров.

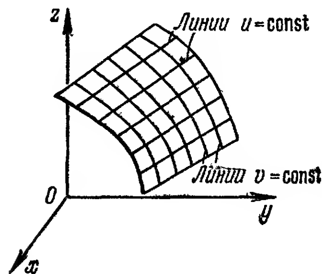


Рис. 211.

Для обоснования этого предположения выберем какие-либо два из уравнений (4), например два первых. Из них можно вообще говоря (во всяком случае, в принципе), выразить u и v через x и y. Если теперь подставить полученные выражения $u=u(x, y)$, $v=v(x, y)$ в третье уравнение (4), то мы получим уравнение вида $z=z(x, y)$, которое, как мы знаем (п. IX.1), представляет поверхность в пространстве. Итак, *уравнения* (4) *задают в параметрической форме поверхность в пространстве.*

В п. XI.14 мы увидим, что в некоторых особых случаях описанный переход принципиально невозможен. Это означает, что рассматриваемая поверхность вырождается в линию или точку. Например, «поверхность» $x=u+v$, $y=2u+2v$, $z=1-u-v$ на самом деле является линией (почему?).

На рис. 211 показан примерный вид поверхности (S), заданной уравнениями (4). Если положить u постоянным и менять только v, то (так как останется только один параметр) мы будем получать различные линии на (S) в зависимости от значения u; аналогично,

полагая $v=\text{const}$, получим другое семейство линий на (S). Эти линии можно принять за координатные линии на (S), а u и v — за координаты на (S).

Уравнения (4) задают определенную функциональную зависимость между x, y, z. Действительно, если задать x и y, то, как говорилось выше, можно (во всяком случае, принципиально) найти соответствующие значения u и v и тем самым соответствующее значение z. Таким образом, получается функция $z(x, y)$, *заданная в параметрической форме*, а графиком ее служит рассмотренная выше поверхность (S).

В общем случае для любого числа переменных *параметрическая зависимость* получается так. Пусть даны уравнения

$$\left.\begin{aligned} x_1&=f_1(t_1, t_2, \ldots, t_m);\\ x_2&=f_2(t_1, t_2, \ldots, t_m);\\ &\ldots\ldots\ldots\ldots\\ x_n&=f_n(t_1, t_2, \ldots, t_m), \end{aligned}\right\} \tag{5}$$

где переменные t_1, t_2, ..., t_m играют роль параметров. Тогда, если $m<n$, то, выбрав из данных уравнений некоторые m, можно (за исключением особых случаев вырождения, о которых будет сказано в п. XI.14) выразить из них t_1, t_2, ..., t_m через соответствующие x, после чего подставить полученные зависимости в остальные уравнения (5). Таким образом, уравнения (5) определяют $n-m$ из x как функции m остальных x. Можно сказать, что уравнения (5) представляют *m-мерное многообразие* (п. 2) в n-мерном пространстве, при выборе точки на таком многообразии имеется m степеней свободы. В случаях вырождения размерность многообразия оказывается меньше m. Если $m\geqslant n$, то уравнения (5) функциональной зависимости между величинами x, вообще говоря, не определяют.

Вычисление производных от функций, заданных параметрически, производится аналогично тому, как в п. IX.13 мы дифференцировали неявные функции. Пусть, например, рассматривается функция $z=z(x, y)$, определенная формулами (4), и мы хотим найти производную z'_x. Тогда, переписав первые два равенства в виде $\varphi(u, v)-x=0$, $\psi(u, v)-y=0$ (при практическом вычислении этого можно и не делать), можно найти u'_x и v'_x, как мы это делали для уравнений (IX.17); условием возможности этого будет

$$\begin{vmatrix} \varphi'_u & \varphi'_v \\ \psi'_u & \psi'_v \end{vmatrix} \neq 0.$$

После этого из последнего равенства (4) имеем формулу $z'_x=\chi'_u u'_x+\chi'_v v'_x$, в которую и нужно подставить найденные значения u'_x и v'_x. Отсюда с помощью дальнейшего дифференцирования можно найти производные высших порядков.

§ 3. Алгебраические поверхности первых двух порядков

7. Поверхности первого порядка. Уравнение поверхности первого порядка в общем виде таково (ср. п. II.9):

$$Ax+By+Cz+D=0. \tag{6}$$

Чтобы выяснить, что это за поверхность, введем вектор

$$\mathbf{a} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}; \tag{7}$$

тогда уравнение (6) можно переписать в форме $\mathbf{a}\cdot\mathbf{r} + D = 0$ [см. формулы (VII.7) и (VII.12)], где $\mathbf{r}$ — радиус-вектор. Однако $\mathbf{a}\cdot\mathbf{r} = a\,\text{пр}_{\mathbf{a}}\,\mathbf{r}$ [см. формулу (VII.4)], откуда получаем

$$a\,\text{пр}_{\mathbf{a}}\,\mathbf{r} + D = 0; \quad \text{пр}_{\mathbf{a}}\,\mathbf{r} = -\frac{D}{a}.$$

Таким образом, получается, что в пространстве нужно взять совокупность всех точек M, для которых проекция радиуса-вектора на постоянный вектор $\mathbf{a}$ имеет постоянное значение $-\frac{D}{a}$. Получится (рис. 212) плоскость (P), перпендикулярная к вектору $\mathbf{a}$.

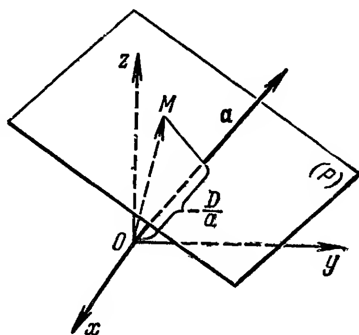


Рис. 212.

Итак, *поверхности первого порядка — это плоскости.*

Рассмотрим несколько простых задач. 1) *Влияние коэффициентов A, B, C, D на положение плоскости (P) в пространстве* видно из рис. 212. Например, если не менять коэффициенты A, B, C, а менять D, то плоскость будет поступательно перемещаться, в частности, при $D = 0$ она пройдет через начало координат; изменение же коэффициентов A, B или C влечет за собой поворот вектора $\mathbf{a}$, а следовательно, и поворот плоскости (P). Если $A = 0$, то вектор $\mathbf{a}$ лежит в плоскости yOz и потому плоскость (P) параллельна оси x; если к тому же и $D = 0$, то плоскость (P) пройдет через ось x.

Аналогично разбираются случаи равенства нулю других коэффициентов.

Отметим уравнения плоскостей xOy: $z = 0$; yOz: $x = 0$ и zOx: $y = 0$.

2) *Через данную точку $(x_1;\ y_1;\ z_1)$ провести плоскость перпендикулярно к данному вектору* (7). Из (6) аналогично первой задаче п. II.9 получаем ответ

$$A(x - x_1) + B(y - y_1) + C(z - z_1) = 0.$$

3) *Угол φ между двумя плоскостями*

$$A_1x + B_1y + C_1z + D_1 = 0, \quad A_2x + B_2y + C_2z + D_2 = 0 \tag{8}$$

равен углу между перпендикулярными к ним векторами

$$\mathbf{a} = A_1\mathbf{i} + B\,\mathbf{j} + C\,\mathbf{k}, \quad \mathbf{a}_2 = A_2\mathbf{i} + B_2\mathbf{j} + C_2\mathbf{k} \tag{9}$$

или дополняет этот последний угол до 180°, как углы с взаимно перпендикулярными сторонами (рис. 213). Значит, косинусы этих углов или равны, или различаются лишь знаком. Вычисляя угол между векторами (см. п. VII.10), получим

$$\cos\varphi = \pm\frac{A_1A_2+B_1B_2+C_1C_2}{\sqrt{A_1^2+B_1^2+C_1^2}\sqrt{A_2^2+B_2^2+C_2^2}}.$$

4) *Условие параллельности*

$$\frac{A_1}{A_2}=\frac{B_1}{B_2}=\frac{C_1}{C_2}$$

двух плоскостей (8) также получается из аналогичного условия (см. задачу 2. п. VII.10) для соответствующих векторов (9). Если же

$$\frac{A_1}{A_2}=\frac{B_1}{B_2}=\frac{C_1}{C_2}=\frac{D_1}{D_2},$$

то уравнения (8) равносильны, т. е. плоскости совпадают.

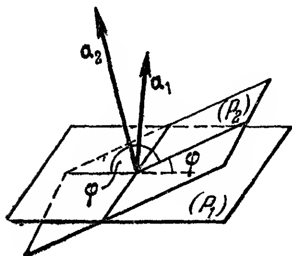


Рис. 213.

5) *Линия* (прямая) *пересечения двух плоскостей* (8) получится, если рассмотреть оба уравнения (8) совместно как систему. От этой системы можно перейти к параметрической форме (VII.33) уравнений прямой так, как это описано в п. 5.

Покажем этот переход на примере. Пусть прямая линия задана как пересечение плоскостей

$$\left.\begin{aligned} x-2y+z-3&=0;\\ 2x+y+4z-5&=0. \end{aligned}\right\}$$

Обозначив $z=t$, получим

$$\left.\begin{aligned} x-2y&=-t+3;\\ 2x+y&=-4t+5. \end{aligned}\right\}$$

Решая эту систему уравнений, найдем $x=\frac{13}{5}-\frac{9}{5}t;\ y=-\frac{1}{5}-\frac{2}{5}t;\ \ z=t.$

Значит (см. рис. 173), рассматриваемая прямая проходит через точку $\left(\frac{13}{5};\ -\frac{1}{5};\ 0\right)$ параллельно вектору $\mathbf{b}=-\frac{9}{5}\,\mathbf{i}-\frac{2}{5}\mathbf{j}+\mathbf{k}$.

Задачи на прямые, а также на плоскости и прямые можно часто решать при помощи такого перехода и применения свойств векторов.

Теперь легко разобраться в *геометрическом смысле различных случаев, которые могут представиться при решении системы* (VI.5) *трех уравнений первой степени с тремя неизвестными*. Каждое из этих уравнений представляет плоскость в пространстве x, y, z, и, таким образом, речь идет об отыскании точки пересечения трех плоскостей (P_1), (P_2), (P_3). Определитель системы D равен векторно-скалярному произведению трех перпендикулярных к ним векторов (п. VII.15). Если $D\neq 0$, то эти векторы не параллельны одной плоскости и поэтому плоскость (P_3) пересечет линию пересечения (P_1) с (P_2)

ровно в одной точке, т. е. система (VI.5) имеет ровно одно решение. Если же $D=0$, то векторы параллельны одной плоскости (T) или, что то же (рис. 214), плоскости (P_1), (P_2), (P_3) параллельны некоторой прямой (l) и поэтому либо не имеют ни одной общей точки, либо, имея общую точку, имеют целую общую прямую, параллельную (l). В первом случае система (VI.5) не имеет решений, а во втором имеет бесконечное количество, целую «прямую решений». Возможное расположение плоскостей в обоих случаях показано на рис. 215. (Какие еще могут быть расположения?)

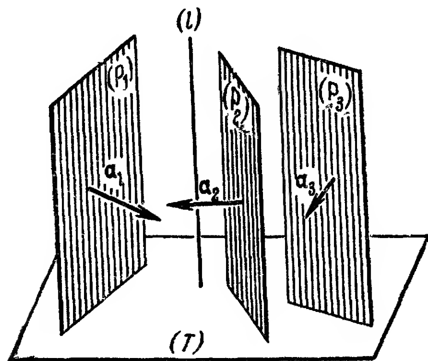


Рис. 214.

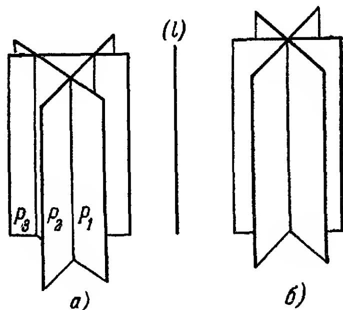


Рис. 215.
а) Общей точки нет; *б*) общая прямая.

8. Эллипсоид. Мы не будем давать геометрическое определение эллипсоида, а начнем с рассмотрения его *канонического уравнения*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, \qquad (10)$$

где a, b, c — положительные постоянные, называемые *полуосями*.

Аналогично п. II.10, легко проверить, что $|x|\leqslant a$, $|y|\leqslant b$, $|z|\leqslant c$, т. е. эллипсоид — конечная, ограниченная поверхность; что плоскости xOy, yOz и zOx служат плоскостями симметрии, а начало координат — центром симметрии (*центр эллипсоида*).

Чтобы уточнить форму эллипсоида, применим *метод сечений*, состоящий в исследовании линий пересечения рассматриваемой поверхности с координатными плоскостями, т. е. плоскостями вида $x=$ $=\text{const}$, $y=\text{const}$ и $z=\text{const}$. Рассмотрим сначала линию пересечения нашего эллипсоида с плоскостью $z=h$, параллельной плоскости xOy. Для этого положим $z=h$ в уравнении (10), в результате чего получится

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1-\frac{h^2}{c^2}, \text{ или } \frac{x^2}{\left(a\sqrt{1-\frac{h^2}{c^2}}\right)^2}+\frac{y^2}{\left(b\sqrt{1-\frac{h^2}{c^2}}\right)^2}=1,$$

т. е. получается эллипс с полуосями $a\sqrt{1-\frac{h^2}{c^2}}$ и $b\sqrt{1-\frac{h^2}{c^2}}$. Таким образом, при $h=0$ получится эллипс с полуосями a и b; при увеличении $|h|$ этот эллипс будет уменьшаться подобно и при

$h=\pm c$ полуоси станут равными нулю, т. е. эллипс выродится в точку. Пересечение эллипсоида с плоскостями $y=h$ и $x=h$ дает аналогичный результат, и мы получаем поверхность, изображенную на рис. 216.

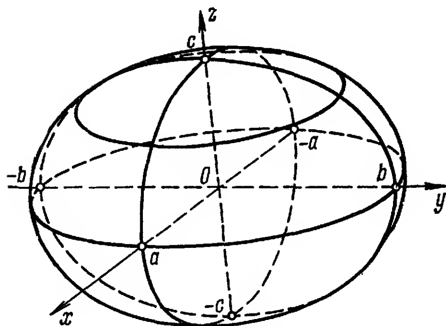


Рис. 216.

Если две из полуосей равны, например $a=b$, то в сечении с плоскостями $z=h$ будут получаться окружности. Тогда вместо *трехосного эллипсоида* получим *эллипсоид вращения*, т. е. поверхность, полученную от вращения эллипса вокруг одной из его осей. В зависимости от того, вокруг какой оси, большой или малой, будет производиться вращение, получится *вытянутый* (как яйцо) или *сплюснутый* эллипсоид вращения. Если же все три полуоси равны, то эллипсоид превращается в *сферу*, т. е. поверхность шара.

Аналогично п. II.10 легко показать, что трехосный эллипсоид получается в результате двух равномерных сжатий или растяжений сферы к координатным плоскостям, в результате одного сжатия получается обязательно эллипсоид вращения. В то же время из п. XI.6 вытекает, что при равномерном сжатии эллипсоида вновь получается эллипсоид.

9. Гиперболоиды. Гиперболоиды бывают двух типов. *Однополостный гиперболоид* имеет каноническое уравнение

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1. \qquad (11)$$

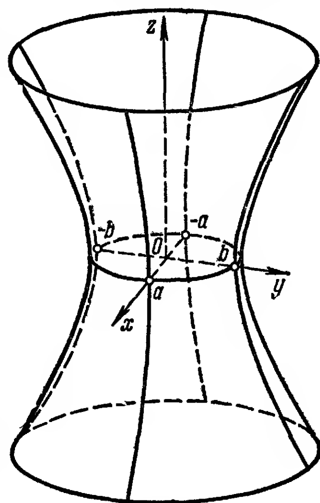


Рис. 217.

Сечение плоскостью $z=h$ дает эллипс с полуосями $a\sqrt{1+\frac{h^2}{c^2}}$ и $b\sqrt{1+\frac{h^2}{c^2}}$ (проверьте!). Значит, при $h=0$ получится эллипс с полуосями a и b, а при увеличении $|h|$ этот эллипс будет подобно увеличиваться до бесконечности. Пересечение с плоскостями $y=h$ и $x=h$ дает гиперболы, и мы получаем поверхность, изображенную на рис. 217. Как и эллипсоид, гиперболоид имеет три плоскости симметрии и центр симметрии.

Если $a=b$, то получается *однополостный гиперболоид вращения* — поверхность, полученная от вращения гиперболы вокруг ее мнимой оси.

Его можно получить также другим способом, для объяснения которого положим $y=b$ в уравнении (11), т. е. посмотрим, что получится в пересечении с плоскостью $y=b$. Получится

$$\frac{x^2}{a^2}-\frac{z^2}{c^2}=0;\quad \left(\frac{x}{a}-\frac{z}{c}\right)\left(\frac{x}{a}+\frac{z}{c}\right)=0,$$

т. е. линия распадается на пару прямых

$$\frac{x}{a}-\frac{z}{c}=0 \quad \text{и} \quad \frac{x}{a}+\frac{z}{c}=0 \qquad (y=b),$$

показанных на рис. 218 и пересекающихся в точке $A\,(0;\ b,\ 0)$. В силу осевой симметрии аналогичная картина получится в пересечении с любой вертикальной плоскостью, касающейся «горловой окружности». Значит, весь *этот гиперболоид целиком заполнен* показанными на рис. 218 *двумя семействами прямых линий*, так что через каждую его точку проходят по две прямые, целиком лежащие на нем (подобно тому как цилиндр или конус целиком заполнены одним семейством прямых линий). Заодно видно, что *однополостный гиперболоид вращения можно получить от вращения одной из двух скрещивающихся прямых в пространстве вокруг другой*. Что касается однополостного гиперболоида общего вида (11), то он получается из гиперболоида вращения в результате равномерного сжатия, при котором прямые остаются прямыми, и поэтому *также заполнен двумя семействами прямых линий*. Отметим в заключение, что плоскость, показанная на рис. 218, является касательной плоскостью к гиперболоиду в точке A. В самом деле, *касательная плоскость в любой точке A любой поверхности* (S) — *это*, по определению, *плоскость, проходящая через A и касательная там к любой линии, проходящей через эту точку и лежащей на* (S). Значит, касательная плоскость к гиперболоиду должна проходить через обе прямые, лежащие на нем и проходящие через A. Итак, оказывается, что касательная плоскость может пересекать поверхность по двум линиям.

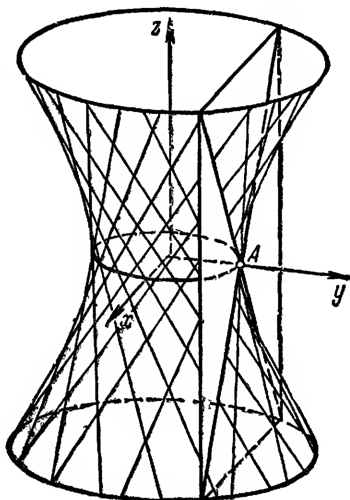


Рис. 218.

Каноническое уравнение *двуполостного гиперболоида* таково:

$$-\frac{x^2}{a^2}-\frac{y^2}{b^2}+\frac{z^2}{c^2}=1.$$

Сечение плоскостью $z=h$ дает эллипс с полуосями

$$a\sqrt{\frac{h^2}{c^2}-1} \quad \text{и} \quad b\sqrt{\frac{h^2}{c^2}-1}.$$

Таким образом, при $|h|<c$ сечения нет, при $|h|=c$ получаются нулевые полуоси, а при дальнейшем увеличении $|h|$ — эллипс, подобно увеличивающийся до бесконечности. Пересечение с плоскостями $y=h$ и $x=h$ дает гиперболы, и мы получаем поверхность, состоящую из

двух бесконечных кусков и изображенную на рис. 219. Если $a=b$, получается *двуполостный гиперболоид вращения* — поверхность, полученная от вращения гиперболы вокруг ее вещественной оси.

Рис. 219.

10. Параболоиды. Параболоиды также бывают двух типов. *Эллиптический параболоид* имеет каноническое уравнение

$$z=ax^2+by^2 \qquad (a,\ b>0).$$

В пересечении с плоскостью $z=h$ получится $ax^2+by^2=h$ или $\frac{x^2}{\frac{h}{a}}+\frac{y^2}{\frac{h}{b}}=1$, т. е. эллипс с полуосями $\sqrt{\frac{h}{a}}$ и $\sqrt{\frac{h}{b}}$. Значит, при $h<0$ пересечения не будет, при $h=0$ получится точка (начало координат), а при увеличении $h>0$ получится эллипс, подобно увеличивающийся до бесконечности. Пересечение с плоскостями $y=h$ и $x=h$ дает параболы, и мы получаем поверхность, изображенную на рис. 220. Параболоид имеет две плоскости симметрии ($x=0$ и $y=0$). Если $a=b$, получается *параболоид вращения* — поверхность, полученная от вращения параболы вокруг ее оси.

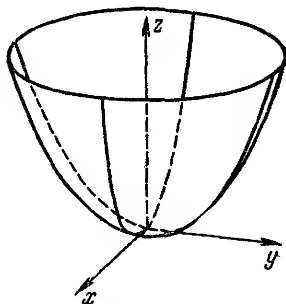


Рис. 220.

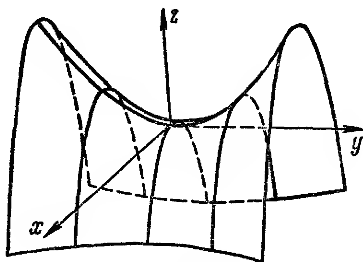


Рис. 221.

Гиперболический параболоид имеет каноническое уравнение

$$z=-ax^2+by^2 \quad (a,\ b>0). \tag{12}$$

Пересечение с плоскостью $x=0$ дает параболу $z=by^2$, обращенную ветвями вверх, тогда как пересечение с плоскостями $y=h$ дает параболы $z=-ax^2+bh^2$, обращенные ветвями вниз (рис. 221). Пересечение же с плоскостями $z=h$ дает гиперболы. Таким образом, получается поверхность, имеющая вид седла.

Можно доказать, что эта поверхность, как и однополостный гиперболоид, целиком заполнена двумя семействами прямых линий. Например, в начале координат касательной плоскостью к поверхности служит плоскость $z=0$; в то же время, если в уравнении (12) положить $z=0$, то мы получим $\sqrt{a}\,x=\pm\sqrt{b}\,y$, т. е. касательная плоскость пересекает поверхность по двум прямым линиям.

11. Обзор поверхностей второго порядка. Уравнение поверхности второго порядка в общем виде (ср. п. II.13) таково:

$$Ax^2+2Bxy+Cy^2+2Dxz+2Eyz+Fz^2+Gx+Hy+Iz+J=0. \qquad (13)$$

В п. XI.12 будет показано, что всегда можно совершить такой поворот декартовых осей координат, после которого уравнение в новых осях уже не будет содержать членов с произведениями координат, т. е. будет иметь вид

$$A'x'^2+C'y'^2+F'z'^2+G'x'+H'y'+I'z'+J=0 \qquad (14)$$

(ср. п. II.13). Дальнейшее исследование идет по-разному, в зависимости от знаков коэффициентов A', C', F'. Допустим сначала, что все эти коэффициенты отличны от нуля и одного знака, например положительны. Тогда, как в п. II.13, производя дополнение до полного квадрата, а затем параллельный перенос осей, приходим к уравнению

$$A'x''^2+C'y''^2+F'z''^2+J'=0, \text{ т. е. } \frac{x''^2}{-\frac{J'}{A'}}+\frac{y''^2}{-\frac{J'}{C'}}+\frac{z''^2}{-\frac{J'}{F'}}=1.$$

Отсюда, если $J'<0$, получаем каноническое уравнение эллипсоида, т. е. и исходная поверхность (13) была эллипсоидом, но относительно осей x, y, z смещенным и повернутым. Если же $J'>0$ или $J'=0$, то получаем соответственно мнимую поверхность или точку. Аналогичные результаты получаются, если коэффициенты A', C', F' отрицательны.

Мы предоставляем читателю убедиться в том, что если коэффициенты A', C', F' отличны от нуля, но имеют разные знаки, то поверхность будет гиперболоидом, либо конусом второго порядка. Этот конус является одновременно эллиптическим, гиперболическим и параболическим (см. рис. 86), но обычно его называют эллиптическим; в частности, может получиться круговой конус.

Если из коэффициентов A', C', F' уравнения (14) ровно один равен нулю, например F', а соответствующий коэффициент I' отличен от нуля, то легко проверить, что поверхность будет параболоидом. Можно проверить, на чем мы не будем останавливаться, что во всех остальных случаях могут получиться лишь цилиндр второго порядка, пара плоскостей (которые могут оказаться совпадающими), вырождение в прямую линию и мнимая поверхность. При этом цилиндр второго порядка может быть эллиптическим (в частности, круговым), гиперболическим или параболическим в зависимости от вида направляющей; например, в начале п. 4 описан параболический цилиндр.

МАТРИЦЫ И ИХ ПРИМЕНЕНИЕ

§ 1. Матрицы

Матрицы, впервые появившиеся в середине прошлого века в работах английских математиков У. Гáмильтона (1805—1865) и А. Кэ́ли (1821—1895), в настоящее время в прикладной математике используются весьма широко, они значительно упрощают рассмотрение сложных систем уравнений.

1. Определения. Мы начнем с формальных определений, целесообразность которых станет постепенно ясной. *Матрицей* называется прямоугольная таблица, составленная из чисел или каких-либо других объектов. Мы будем, если не оговорено противное, рассматривать только *вещественные числовые матрицы*, т. е. матрицы, составленные из вещественных чисел. Такая матрица может иметь, например, вид

$$\begin{pmatrix} 2 & -1{,}3 & 0 \\ 1 & \pi & 1 \end{pmatrix}, \text{ или } \begin{pmatrix} 2 & 1 & 0 \\ -3 & \sqrt{2} & 0 \\ 2 & -1 & {}^1/_2 \end{pmatrix}, \text{ или } \begin{pmatrix} 1 \\ -2 \\ 0 \\ 3 \end{pmatrix}, \text{ или } (5) \text{ и т. д.} \qquad (1)$$

Здесь круглые скобки по бокам — *знак матрицы*; применяются также волнистые вертикальные линии, но, конечно, не простые вертикальные черточки, которыми обозначается определитель (§ VI.1). Как у определителей, у матриц различают элементы, строчки и столбцы. Однако важнейшим отличием от определителей является то, что определитель считается равным некоторому числу (п. VI.1), тогда как матрица не приравнивается какому-либо более простому объекту. Ее можно для краткости обозначить одной буквой, например **A**, **B** и т. п., но тогда под **A** все равно будет пониматься таблица. В общем виде

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \qquad (2)$$

Таким образом, элементы матрицы удобно снабжать двумя индексами, из которых первый указывает номер строки, а второй — номер столбца. Иногда коротко пишут $A = (a_{ij})_{mn}$, т. е. i меняется от 1 до m, а j — от 1 до n.

Каждая матрица имеет определенные *размеры*, т. е. количество строк и количество столбцов; так, в строках (1) и (2) выписаны матрицы соответственно размера 2×3, 3×3, 4×1, 1×1, $m\times n$. Если число строк равно числу столбцов, то матрица называется *квадратной*, тогда говорят о ее *порядке*. Квадратная матрица первого порядка отождествляется со своим единственным элементом; так, четвертая матрица (1) — это просто число 5.

Матрица, у которой всего один столбец, называется *столбцевой*, или *числовым вектором*; такая матрица отождествляется с вектором в вещественном числовом пространстве (п. VII.18). Так, третья матрица (1) — это вектор в E_4 с координатами 1; —2; 0; 3. Матрица, у которой всего одна строка, называется *строчной*.

Матрица, у которой все элементы равны нулю, называется *нулевой*. Квадратная матрица, у которой равны нулю все элементы, кроме, быть может, стоящих на *главной диагонали* (т. е. диагонали, идущей из левого верхнего в правый нижний угол), называется *диагональной*; если на диагонали стоят элементы a, b, $\ldots$, k, то матрица обозначается $\operatorname{diag}(a, b, \ldots, k)$. Диагональная матрица, у которой все диагональные элементы равны единице, называется *единичной* и обычно обозначается буквой $\mathbf{I}$: например, единичная матрица третьего порядка имеет вид

$$\mathbf{I}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}=\operatorname{diag}(1,\ 1,\ 1). \tag{3}$$

Иногда применяется *транспонирование* матрицы $\mathbf{A}$, т. е. перемена ролями ее строк и столбцов (ср. п. VI.2); полученную матрицу мы обозначим $\mathbf{A}^*$. Например,

$$\begin{pmatrix}2 & 0 & 3\\ 7 & -2 & 1\end{pmatrix}^*=\begin{pmatrix}2 & 7\\ 0 & -2\\ 3 & 1\end{pmatrix},\quad \begin{pmatrix}1\\ 0\\ -2\end{pmatrix}^*=(1\ \ 0\ \ -2);$$

в общем виде можно написать $a^*_{ij}=a_{ji}$ (почему?). Ясно, что $(\mathbf{A}^*)^*=\mathbf{A}$.

Матрица, совпадающая со своей транспонированной, называется *симметрической*; такой может быть только квадратная матрица. Условие симметричности можно записать в виде $a_{ij}=a_{ji}$.

Если $a_{ij}=-a_{ji}$, то матрица называется *кососимметрической*.

Квадратная матрица $\mathbf{A}$ *имеет определитель*, который мы будем обозначать $\det\mathbf{A}$; например, $\det\begin{pmatrix}1 & 0\\ 2 & -3\end{pmatrix}=\begin{vmatrix}1 & 0\\ 2 & -3\end{vmatrix}=-3$. Прямоугольная неквадратная матрица определителя не имеет, так как *определители бывают только квадратные*. Из п. VI.2 следует, что

$$\det\mathbf{I}=1,\quad \det\mathbf{A}^*=\det\mathbf{A}.$$

2. Действия над матрицами. Матрицы *одинакового размера* можно *складывать* по формуле

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{pmatrix};$$

аналогично определяется *умножение матрицы на число:*

$$k\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} ka_{11} & ka_{12} & ka_{13} \\ ka_{21} & ka_{22} & ka_{23} \end{pmatrix}.$$

Легко проверить, что при этом выполняются все аксиомы линейных действий (п. VII.17), т. е. совокупность всех матриц одинакового размера образует линейное пространство. Отметим очевидные формулы:

$$(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*; \quad (k\mathbf{A})^* = k\mathbf{A}^*; \ \det(k\mathbf{C}) = k^n \det \mathbf{C},$$

где n — порядок квадратной матрицы $\mathbf{C}$. При этом, вообще говоря, $\det(\mathbf{A}+\mathbf{B}) \neq \det \mathbf{A} + \det \mathbf{B}$.

Умножение матриц друг на друга осуществляется по своеобразному закону, разумность которого будет ясна из п. 6. Прежде всего, размеры матриц-сомножителей должны быть согласованы: ширина первого множителя должна равняться высоте второго, в противном случае умножение невозможно. Если же это условие выполнено, то произведение находится по следующему правилу:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} & a_{11}b_{13}+a_{12}b_{23} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} & a_{21}b_{13}+a_{22}b_{23} \end{pmatrix}.$$

Надо внимательно продумать это правило. Например, чтобы получить у произведения элемент, стоящий в первой строке и в третьем столбце, надо у первого множителя взять первую строку, у второго — третий столбец, а затем эти строку и столбец как бы скалярно перемножить (см. формулу (VII.12)). И другие элементы матрицы-произведения получаются с помощью аналогичного «как бы скалярного умножения» *строк первой матрицы*-множителя *на столбцы второй матрицы*-множителя. В общем случае, если мы умножим матрицу (a_{ij}) размера $m \times n$ на матрицу (b_{ij}) размера $n \times p$, мы получим матрицу (c_{ij}) размера $m \times p$, элементы которой вычисляются по формуле $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$.

Из приведенного правила вытекает, что всегда возможно перемножить две квадратные матрицы одинакового порядка, что даст квадратную матрицу того же порядка. В частности, квадратную матрицу всегда можно умножить саму на себя, т. е. возвысить в квадрат, тогда как прямоугольную неквадратную матрицу возвысить в квадрат нельзя. Другим важным частным случаем является умножение строч-

ной матрицы на столбцевую, причем ширина первой равна высоте второй; это даст квадратную матрицу первого порядка, т. е. число

$$(a_1\ a_2\ a_3)\cdot\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1b_1 + a_2b_2 + a_3b_3.$$

Аналогично п. VI.2 можно проверить следующие свойства произведения матриц:

$$(k\mathbf{A})\,\mathbf{B} = \mathbf{A}\,(k\mathbf{B}) = k\,(\mathbf{AB});\quad (\mathbf{A}+\mathbf{B})\,\mathbf{C} = \mathbf{AC}+\mathbf{BC};$$
$$\mathbf{C}\,(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB};\quad \mathbf{A}\,(\mathbf{BC}) = (\mathbf{AB})\,\mathbf{C};$$

конечно, при этом всегда подразумевается, что размеры участвующих здесь матриц обеспечивают осмысленность формулы. Другой метод вывода этих формул будет указан в п. 6.

Уже на самых простых примерах легко проверить, что *матрицы, вообще говоря, не перестановочны друг с другом*, т. е., вообще говоря, $\mathbf{AB} \neq \mathbf{BA}$. Например (проверьте!)

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\cdot\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},\quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\cdot\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix};$$

$$(1\ \ 2)\begin{pmatrix} 0 \\ -2 \end{pmatrix} = -4,\quad \begin{pmatrix} 0 \\ -2 \end{pmatrix}(1\ \ 2) = \begin{pmatrix} 0 & 0 \\ -2 & -4 \end{pmatrix};$$

$$\begin{pmatrix} 1 & -2 \\ -1 & -3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ -7 \end{pmatrix},\quad \begin{pmatrix} 1 \\ 2 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -1 & -3 \end{pmatrix} \text{ не имеет смысла.}$$

Поэтому при умножении матриц надо тщательно следить за порядком множителей; для этого применяются термины: «умножим **A** *справа* на **B**» или просто «умножим **A** на **B**» (получится **AB**), но «умножим **A** *слева* на **B**» (получится **BA**).

Отметим еще непосредственно проверяемое свойство

$$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^* \tag{4}$$

и свойство, которое будет доказано в п. 7,

$$\det(\mathbf{AB}) = \det\mathbf{A}\cdot\det\mathbf{B}. \tag{5}$$

Если **A** — *комплексная числовая матрица*, то под $\mathbf{A}^*$ понимается результат транспонирования с одновременной заменой всех элементов на их комплексно сопряженные значения; при этом $\mathbf{A}^*$ называется матрицей, *сопряженной* с **A**. Для комплексных матриц из приведенных выше формул надо изменить только две: $\det\mathbf{A}^* = (\det\mathbf{A})^*$, $(k\mathbf{A})^* = k^*\mathbf{A}^*$.

3. Обратная матрица. Будем рассматривать *квадратные* матрицы некоторого определенного, например третьего, порядка. При умножении таких матриц единичная матрица (3) играет ту же роль, что единица при умножении чисел: легко непосредственно проверить, что $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ для любой матрицы **A**.

По аналогии с умножением чисел определяется и понятие *обратной к* $\mathbf{A}$ матрицы: это матрица $\mathbf{A}^{-1}$, для которой

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}. \tag{6}$$

Отсюда и из равенства (5) вытекает, что

$$\det \mathbf{A} \cdot \det(\mathbf{A}^{-1}) = \det \mathbf{I} = 1, \quad \text{т. е.} \quad \det(\mathbf{A}^{-1}) = \frac{1}{\det \mathbf{A}}.$$

Мы видим, что обязательно должно быть $\det \mathbf{A} \neq 0$. Квадратная матрица $\mathbf{A}$, для которой $\det \mathbf{A} = 0$, называется *вырожденной*. Таким образом, *вырожденная матрица не имеет обратной*. В то же время *всякая невырожденная матрица имеет обратную*. В самом деле, рассмотрим любую невырожденную матрицу

$$\mathbf{K} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}. \tag{7}$$

Тогда, исходя из определения произведения матриц и рассуждая, как в п. VI.4, легко проверить, что произведение $\mathbf{K}$ слева или справа на матрицу

$$\frac{1}{\det \mathbf{K}} \begin{pmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{pmatrix} \tag{8}$$

равно $\mathbf{I}$; здесь большими буквами $A_1, \ldots, C_3$ обозначены алгебраические дополнения соответствующих элементов в матрице $\mathbf{K}$ (или, что то же, в соответствующем определителе, см. п. VI.3). Значит, матрица (8) и есть $\mathbf{K}^{-1}$.

Обратные матрицы применяются при решении матричных уравнений. Например, рассмотрим уравнение $\mathbf{AX} = \mathbf{B}$, где $\mathbf{A}$ и $\mathbf{B}$ — заданные матрицы, а $\mathbf{X}$ — искомая, причем $\det \mathbf{A} \neq 0$. Умножив обе части слева на $\mathbf{A}^{-1}$ и воспользовавшись равенствами (6), получаем $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$. Аналогично из уравнения $\mathbf{XA} = \mathbf{B}$ получаем решение $\mathbf{X} = \mathbf{B}\mathbf{A}^{-1}$.

Матрицы дают возможность кратко записать систему уравнений первой степени. Например, систему уравнений (VI.5) можно переписать в *матричной форме*

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}$$

(проверьте!). Если обозначить матрицу коэффициентов буквой $\mathbf{A}$, столбец (т. е. числовой вектор) неизвестных буквой $\mathbf{x}$, а столбец свободных членов буквой $\mathbf{d}$, то то же уравнение можно еще короче записать в виде

$$\mathbf{Ax} = \mathbf{d}, \tag{9}$$

откуда при $\det A \neq 0$ сразу получаем решение

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{d}. \qquad (10)$$

Конечно, если подробно расшифровать эту формулу, получится то же правило Крамера, которое было найдено в п. VI.4. Отметим, что запись (9) возможна и в случае, когда число уравнений не равно числу неизвестных, тогда матрица **A** не будет квадратной; однако при этом нельзя перейти к формуле (10), так как *неквадратная матрица не имеет обратной.*

Из формул (6) видно, что матрицы **A** и $\mathbf{A}^{-1}$ являются взаимно обратными, т. е. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$. Кроме этой, иногда применяется формула $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ ($\det \mathbf{A} \neq 0$, $\det \mathbf{B} \neq 0$), которую легко проверить: $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{I}\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$.

Наконец, из формулы (4), если подставить $\mathbf{B} = \mathbf{A}^{-1}$, вытекает, что $(\mathbf{A}^{-1})^*\mathbf{A}^* = (\mathbf{A}\mathbf{A}^{-1})^* = \mathbf{I}^* = \mathbf{I}$, т. е. $(\mathbf{A}^{-1})^* = (\mathbf{A}^*)^{-1}$.

4. Собственные векторы и собственные значения матрицы. Пусть **A** — заданная *квадратная* матрица. Как мы увидим позже, иногда приходится рассматривать уравнение

$$\mathbf{Ax} = \lambda \mathbf{x}, \qquad (11)$$

где **x** — неизвестный числовой вектор, высота которого равна порядку **A**, а λ — неизвестное число. При любом λ уравнение (11) обладает, в частности, *тривиальным решением* $\mathbf{x} = \mathbf{0}$, однако нас будут интересовать только такие λ, при которых эта система имеет *нетривиальные решения.* Эти значения λ называются *собственными значениями* матрицы **A**, а решения **x** уравнения (11) при таких λ — ее *собственными векторами.*

Собственные значения и собственные векторы находятся следующим образом. Так как $\mathbf{x} = \mathbf{Ix}$, то уравнение (11) можно переписать в виде

$$(\mathbf{A} - \lambda \mathbf{I})\,\mathbf{x} = \mathbf{0}. \qquad (12)$$

Сравнивая с формулой (9), видим, что получилась система из n алгебраических линейных однородных уравнений с n неизвестными, где n — порядок матрицы **A**. Согласно п. VI.6 для наличия нетривиального решения необходимо и достаточно, чтобы определитель системы равнялся нулю, т. е.

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0. \qquad (13)$$

Это уравнение называется *характеристическим уравнением* матрицы **A**, оно служит для разыскания собственных значений λ. Так, для матрицы (7) оно имеет вид

$$\begin{vmatrix} a_1 - \lambda & b_1 & c_1 \\ a_2 & b_2 - \lambda & c_2 \\ a_3 & b_3 & c_3 - \lambda \end{vmatrix} = 0.$$

Раскрыв определитель, мы видим, что получается алгебраическое уравнение, степень которого равна порядку матрицы **A**. В силу п. VIII.8 заключаем, что *матрица порядка n имеет n собственных значений*, среди которых, правда, могут быть совпадающие.

Найдя какое-либо собственное значение, мы можем соответствующие собственные векторы найти из векторного уравнения (12) (переписанного в виде системы скалярных уравнений), как указано в п. VI. 6. Из уравнения (12) вытекает, что при зафиксированном λ сумма решений $\mathbf{y}=\mathbf{x}^1+\mathbf{x}^2$ будет снова решением и произведение $\mathbf{y}=k\mathbf{x}$ решения на число будет также решением того же уравнения. Значит, *совокупность всех собственных векторов, отвечающих заданному собственному значению, образует линейное подпространство* (п. VII.18) *пространства всех числовых векторов* заданной высоты n.

В наиболее важном случае, когда все собственные значения различные, каждое из этих подпространств одномерное, т. е. для каждого собственного значения соответствующий собственный вектор определен с точностью до числового множителя. При этом имеются в виду комплексная размерность и комплексные собственные векторы, так как вещественное характеристическое уравнение (13) может иметь как вещественные, так и мнимые корни. Указанная одномерность вытекает из того, что ненулевые собственные векторы, отвечающие различным собственным значениям, обязательно линейно независимы, а в n-мерном пространстве числовых векторов не может быть более n линейно независимых векторов. А эта линейная независимость проверяется так: если, например, собственные векторы $\mathbf{x}^1$, $\mathbf{x}^2$, $\mathbf{x}^3$ отвечают различным собственным значениям λ_1, λ_2, λ_3, причем $\mathbf{x}^1$ с $\mathbf{x}^2$ линейно независимы, а $\mathbf{x}^3=\alpha\mathbf{x}^1+\beta\mathbf{x}^2$, то, помножив это равенство справа на **A**, получаем $\lambda_3\mathbf{x}^3=\alpha\lambda_1\mathbf{x}^1+\beta\lambda_2\mathbf{x}^2$, откуда, умножив первое равенство на λ_3 и вычитая, выводим $\alpha(\lambda_1-\lambda_3)\mathbf{x}^1+\beta(\lambda_2-\lambda_3)\mathbf{x}^2=0$, чему противоречит линейная независимость $\mathbf{x}^1$ и $\mathbf{x}^2$.

Если имеются совпадающие собственные значения, то можно проверить, что для каждого собственного значения λ_k кратности n_k подпространство собственных векторов имеет размерность $m_k\leqslant n_k$. Если все $m_k=n_k$, то, выбрав базис в каждом из этих подпространств, мы получаем базис в комплексном числовом пространстве Z_n, состоящий из собственных векторов матрицы **A**, имеющей порядок n (если все λ_k вещественные, получаем базис в E_n). Если хотя бы одно $m_k<n_k$, то базиса из собственных векторов матрицы **A** указать нельзя.

5. Ранг матрицы. Вычеркнем из матрицы **A** несколько строк и столбцов так, чтобы количество оставшихся строк равнялось количеству оставшихся столбцов. Если после этого заменить знак матрицы на знак определителя, то полученный определитель называется *минором* матрицы **A**. Матрица имеет много миноров, причем некоторые из них могут равняться нулю, а другие отличны от нуля. *Наивысший из порядков миноров, отличных от нуля, называется рангом матрицы* **A**, это очень важная ее характеристика. Например, у матрицы $\mathbf{B}=\begin{pmatrix}2 & 0 & -4\\ 3 & 0 & -6\end{pmatrix}$ все три минора второго порядка $\begin{vmatrix}0 & -4\\ 0 & -6\end{vmatrix}$, $\begin{vmatrix}2 & -4\\ 3 & -6\end{vmatrix}$, $\begin{vmatrix}2 & 0\\ 3 & 0\end{vmatrix}$ равны нулю, тогда как среди шести миноров первого порядка имеется четыре отличных от нуля. (Определитель первого порядка

принимается равным своему единственному элементу.) Поэтому $\operatorname{rang} \mathbf{B} = 1$. Мы предоставляем читателю проверить, что ранги матриц

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 3 & 0 & 2 \\ 1 & -1 & 3 \\ 4 & -1 & 6 \end{pmatrix}, \quad \begin{pmatrix} 3 & 0 & 2 \\ 1 & -1 & 3 \\ 4 & -1 & 5 \end{pmatrix}, \quad \begin{pmatrix} 3 & 0 & 2 \\ 6 & 0 & 4 \\ 9 & 0 & 6 \end{pmatrix}, \quad (3 \ 0 \ 2) \qquad (14)$$

равны соответственно 2, 3, 2, 1, 1, Ранг нулевой матрицы, у которой вовсе нет миноров, отличных от нуля, принимается равным нулю.

Ясно, что ранг квадратной матрицы не превосходит ее порядок; ранг равен порядку в том и только том случае, если матрица невырожденная. Ранг матрицы размера $m \times n$, где $m \neq n$, не превосходит меньшего из чисел m и n.

Можно доказать, на чем мы не будем останавливаться, что *ранг матрицы равен максимально возможному числу ее линейно независимых строк.* (Отметим, что строки матрицы сами являются матрицами, т. е. над ними возможно производить линейные действия.) Так, во втором примере (14) все три строки линейно независимы; в третьем примере первые две строки линейно независимые, а третья равна их сумме; в четвертом примере вторая и третья строки линейно выражаются через первую.

Из свойства 7 п. VI.2 сразу следует, что *у транспонированных матриц ранги одинаковы.* Поэтому *ранг* одновременно *равен максимально возможному числу линейно независимых столбцов матрицы.*

С помощью понятия ранга формулируются окончательные теоремы о разрешимости систем линейных алгебраических уравнений, даже если число уравнений не равно числу неизвестных. Рассмотрим для определенности систему из трех уравнений с четырьмя неизвестными

$$\left.\begin{aligned} a_1x + b_1y + c_1z + d_1u = f_1, \\ a_2x + b_2y + c_2z + d_2u = f_2, \\ a_3x + b_3y + c_3z + d_3u = f_3. \end{aligned}\right\} \qquad (15)$$

Если ввести числовые векторы

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \ \ldots, \ \mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix},$$

то систему (15) можно переписать в виде

$$\mathbf{f} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} + u\mathbf{d}, \qquad (16)$$

т. е. задача сводится к разложению заданного вектора $\mathbf{f}$ по четырем заданным векторам $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$. Когда это возможно? Все векторы вида $x\mathbf{a} + y\mathbf{b} + z\mathbf{c} + u\mathbf{d}$ при заданных $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ и всевозможных x, y, z, u образуют линейное подпространство в E_3, «натянутое» на $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$. Размерность этого подпространства в силу леммы п. VII.19 равна максимальному числу k линейно независимых векторов среди $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, т. е. рангу матрицы $\mathbf{A}$ коэффициентов системы (15). Для разложимости (16) нужно, чтобы вектор $\mathbf{f}$ лежал в указанном подпространстве, т. е. чтобы среди векторов $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, $\mathbf{f}$ было также только k линейно независимых. Итак, получаем необходимое и достаточное условие существования решения системы (15):

$$\operatorname{rang}\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix} = \operatorname{rang}\begin{pmatrix} a_1 & b_1 & c_1 & d_1 & f_1 \\ a_2 & b_2 & c_2 & d_2 & f_2 \\ a_3 & b_3 & c_3 & d_3 & f_3 \end{pmatrix}. \qquad (17)$$

Аналогичный вид имеет условие разрешимости для любого числа уравнений и неизвестных.

Пусть теперь условие разрешимости (17) выполнено; сколько тогда решений имеет система (15)? Если обозначить через x_0, y_0, z_0, u_0 какое-либо одно решение этой системы и ввести замену переменных $x = x_0 + x'$, ..., $u = u_0 + u'$, то легко проверить, что x', y', z', u' должны удовлетворять соответствующей однородной системе

$$\left.\begin{aligned} a_1x' + b_1y' + c_1z' + d_1u' = 0,\\ a_2x' + b_2y' + c_2z' + d_2u' = 0,\\ a_3x' + b_3y' + c_3z' + d_3u' = 0 \end{aligned}\right\} \qquad (18)$$

Введем в E_4 числовые векторы

$$\mathbf{p}_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{pmatrix}, \quad \mathbf{p}_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix}, \quad \mathbf{p}_3 = \begin{pmatrix} a_3 \\ b_3 \\ c_3 \\ d_3 \end{pmatrix}, \quad \mathbf{x}' = \begin{pmatrix} x' \\ y' \\ z' \\ u' \end{pmatrix}.$$

Тогда в силу п. VII.20—21 систему (18) можно переписать в виде

$$\mathbf{p}_1 \cdot \mathbf{x}' = 0, \quad \mathbf{p}_2 \cdot \mathbf{x}' = 0, \quad \mathbf{p}_3 \cdot \mathbf{x}' = 0. \qquad (19)$$

Таким образом, искомый вектор $\mathbf{x}'$ должен быть перпендикулярен к подпространству в E_4, «натянутому» на $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$. Размерность этого подпространства равна рангу (17), а потому нетрудно проверить, что размерность линейного подпространства векторов $\mathbf{x}'$ равна $4 - \operatorname{rang} \mathbf{A}$ (в общем случае вместо 4 должно быть число неизвестных). Такой же получается и размерность совокупности решений системы (19), если каждое решение рассматривать как набор координат точки в E_4, то получается, что при выполнении условия (17) совокупность решений системы (15) определяет в E_4 гиперплоскость (п. VII.19) размерности $4 - \operatorname{rang} \mathbf{A}$.

§ 2. Линейные отображения

6. Линейное отображение и его матрица. Начнем с примера. Пусть плоскость повернута на угол α. Тогда любой вектор $\vec{x}$ в плоскости перейдет в другой вектор $\vec{y}$, который мы обозначим буквами $\mathrm{A}(\vec{x})$ или просто $\mathrm{A}\vec{x}$, т. е. $\vec{y} = \mathrm{A}\vec{x}$. (По причине, которая вскоре станет ясной, мы будем в этом параграфе обозначать векторы светлыми буквами со стрелкой.) Таким образом, в данном случае A — это знак поворота вектора, каждому вектору $\vec{x}$ отвечает вектор $\mathrm{A}\vec{x}$. Иначе говорят, что задано *отображение* A плоскости векторов в себя; употребляются также слова *преобразование* A, *оператор* A. Вектор $\vec{x}$ называется *прообразом*, а вектор $\mathrm{A}\vec{x}$ — *образом* при отображении A.

Так как при повороте параллелограмм переходит в параллелограмм, то при сложении прообразов складываются также и образы (рис. 222), другими словами, образом суммы служит сумма образов, $\mathrm{A}(\vec{x}_1 + \vec{x}_2) = \mathrm{A}\vec{x}_1 + \mathrm{A}\vec{y}_2$. Аналогично проверяется, что при умножении прообраза на число образ умножится на то же число, $\mathrm{A}(\lambda\vec{x}) = \lambda\mathrm{A}\vec{x}$. Таким образом, при рассматриваемом отображении сохраняются ли-

нейные действия, т. е. сохраняются линейные соотношения между векторами; это свойство отображения называется его линейностью.

Аналогичным свойством линейности обладает проектирование всех векторов в пространстве на некоторую фиксированную плоскость или прямую (если при этом проекцию вектора рассматривать как вектор), что мы предоставим доказать читателю. Совсем другой пример получится, если рассмотреть линейное пространство всех многочленов (п. VII.18) и под образом каждого многочлена понимать его производную; линейность такого отображения вытекает из того, что производная суммы равна сумме производных, а постоянный множитель можно выносить за знак производной.

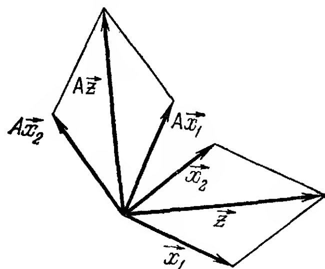


Рис. 222.
$\vec{z}=\vec{x}_1+\vec{x}_2$, $A\vec{z}=A\vec{x}_1+A\vec{x}_2$,
$A(\vec{x}_1+\vec{x}_2)=A\vec{x}_1+A\vec{x}_2$.

Переходим к общему определению. Пусть задано два линейных пространства (п. VII.17), (R) и (S), и пусть каждому вектору $\vec{x}\in(R)$ по какому-то закону отвечает вектор $\vec{y}=A\vec{x}\in(S)$; тогда говорят, что задано *отображение* A *пространства* (R) *в пространство* (S). (Если $(S)=(R)$, то говорят об отображении пространства (R) в *себя*.) Это отображение называется *линейным*, если для любых $\vec{x}_1, \vec{x}_2\in(R)$ и любого числа λ

$$A(\vec{x}_1+\vec{x}_2)=A\vec{x}_1+A\vec{x}_2, \quad A(\lambda\vec{x})=\lambda A\vec{x}. \tag{20}$$

Применяя эти свойства несколько раз, легко вывести, что

$$A(\lambda_1\vec{x}_1+\lambda_2\vec{x}_2+\ldots+\lambda_k\vec{x}_k)=\lambda_1 A\vec{x}_1+\lambda_2 A\vec{x}_2+\ldots+\lambda_k A\vec{x}_k, \tag{21}$$

т. е. линейное отображение сохраняет вид линейной комбинации: не только сумма прообразов переходит в сумму образов, но и разность переходит в разность и т. д. Положив во втором равенстве (20) $\lambda=0$, мы получаем также, что всегда $A\vec{0}=\vec{0}$; конечно, в левой части здесь стоит нуль-вектор пространства (R), а в правой части — нуль-вектор пространства (S).

Допустим для определенности, что пространство (R) трехмерное, а пространство (S) двумерное, и как-то выберем в (R) базис $\vec{p}_1, \vec{p}_2, \vec{p}_3$, а в (S) — базис $\vec{q}_1, \vec{q}_2$. Каждый из векторов $A\vec{p}_j$ лежит в (S) и потому можно разложить по $\vec{q}_1, \vec{q}_2$; обозначим

$$\left.\begin{aligned} A\vec{p}_1&=a_{11}\vec{q}_1+a_{21}\vec{q}_2,\\ A\vec{p}_2&=a_{12}\vec{q}_1+a_{22}\vec{q}_2,\\ A\vec{p}_3&=a_{13}\vec{q}_1+a_{23}\vec{q}_2. \end{aligned}\right\} \tag{22}$$

Тогда в силу формулы (21) любой вектор $\vec{x}=x_1\vec{p}_1+x_2\vec{p}_2+x_3\vec{p}_3\in(R)$ переходит в вектор $\vec{y}=\mathrm{A}\vec{x}=y_1\vec{q}_1+y_2\vec{q}_2\in(S)$ по следующему правилу:

$$\vec{y}=\mathrm{A}(x_1\vec{p}_1+x_2\vec{p}_2+x_3\vec{p}_3)=x_1\mathrm{A}\vec{p}_1+x_2\mathrm{A}\vec{p}_2+x_3\mathrm{A}\vec{p}_3=$$
$$=(a_{11}x_1+a_{12}x_2+a_{13}x_3)\vec{q}_1+(a_{21}x_1+a_{22}x_2+a_{23}x_3)\vec{q}_2,$$

т. е.

$$\left.\begin{aligned} y_1&=a_{11}x_1+a_{12}x_2+a_{13}x_3,\\ y_2&=a_{21}x_1+a_{22}x_2+a_{23}x_3.\end{aligned}\right\}\qquad(23)$$

Мы пришли к формулам, по которым преобразуются координаты вектора при линейном отображении. Если обозначить

$$\mathbf{x}=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix},\qquad \mathbf{y}=\begin{pmatrix}y_1\\y_2\end{pmatrix},\qquad \mathbf{A}=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\end{pmatrix},$$

то формулы (23) можно переписать в силу п. 2 в виде

$$\mathbf{y}=\mathbf{A}\mathbf{x}.\qquad(24)$$

Здесь числовая матрица $\mathbf{A}$ называется *матрицей отображения* A *в выбранных базисах* $\vec{p}_j$ и $\vec{q}_i$, так как она зависит не только от сáмого отображения, но и от выбора указанных базисов; аналогично числовые векторы $\mathbf{x}$ и $\mathbf{y}$ зависят не только от векторов $\vec{x}$, $\vec{y}$, но и от этих базисов.

Обратно, если в пространствах (R) и (S) выбраны базисы соответственно $\vec{p}_j$ и $\vec{q}_i$ и дано, что любой вектор $\vec{x}=x_1\vec{p}_1+x_2\vec{p}_2+x_3\vec{p}_3\in(R)$ переходит в вектор $\vec{y}=\mathrm{A}\vec{x}=y_1\vec{q}_1+y_2\vec{q}_2\in(S)$ в соответствии с формулами (23), то преобразование $\vec{y}=\mathrm{A}\vec{x}$ будет линейным.

В самом деле, при сложении векторов их одноименные координаты складываются, а потому и соответствующие числовые векторы складываются, но из формулы (24) видно, что при сложении числовых векторов $\mathbf{x}$ соответствующие векторы $\mathbf{y}$ также складываются. Аналогично проверяется второе свойство (20).

Итак, *при выбранных базисах* $\vec{p}_j$ *и* $\vec{q}_i$ *каждому линейному отображению* (R) *в* (S) *отвечает его матрица, и обратно, каждая числовая матрица соответствующего размера* (в разбираемом случае — размера 2×3) *служит матрицей линейного отображения* (R) *в* (S). Ясно, что если (R) имеет размерность n, а (S) — размерность m, получаются матрицы размеров $m\times n$. В частном случае, когда $(S)=(R)$, матрица отображения будет квадратной; тогда обычно считается, если не оговорено противное, что как до отображения, так и после векторы раскладываются по одному и тому же базису.

Так как ранг матрицы $\mathbf{A}$ равен максимальному числу ее линейно независимых столбцов, то из формул (22) следует, что этот ранг равен максимальному числу линейно независимых векторов $\mathrm{A}\vec{p}_j$, т. е. размерности линей-

ного пространства А (R) образов всех векторов из (R). Это пространство А (R) может либо совпадать с (S), либо образовывать подпространство (S) меньшей размерности; в первом случае говорят, что (R) отображено *на* (S). Из сказанного, в частности, вытекает, что хотя матрица **А** отображения А зависит от выбора базисов в пространствах (R) и (S), но ранг этой матрицы не зависит от такого выбора. Кроме того, так как ранг матрицы размера $m\times n$ не может быть больше n, то получаем, что размерность А (R) не больше размерности (R), т. е. при линейном отображении (а также, как мы увидим в § 4, и при нелинейном отображении) размерность не может повыситься.

Можно рассматривать не только отображение векторов на векторы, но также отображение точек на точки, что более наглядно. Допустим, что каждой точке M плоскости (P) отвечает некоторая точка $\overline{M}$ плоскости $(\overline{P})$, т. е. задано отображение плоскости (P) в плоскость $(\overline{P})$; мы будем писать $\overline{M}=f(M)$ (ср. по поводу этого обозначения п. IX.9). Пусть это отображение f таково, что прямолинейность при нем не нарушается, т. е. векторы плоскости (P) переходят в векторы плоскости $(\overline{P})$. Кроме того, допустим, что при этом равные векторы плоскости (P) переходят в равные векторы плоскости $(\overline{P})$, т. е. любому вектору $\vec{x}$ плоскости (P), вне зависимости от его расположения, отвечает вполне определенный вектор $\vec{y}$ плоскости $(\overline{P})$, который мы обозначим $A\vec{x}$. Наконец, допустим, что отображение $\vec{y}=A\vec{x}$ линейное. (Пример, с которого мы начали п. 6, удовлетворяет всем перечисленным предположениям, в нем плоскости (P) и $(\overline{P})$ совпадают, а под $f(M)$ надо понимать результат поворота точки M на угол α вокруг центра вращения.)

Пусть теперь в плоскости (P) выбрана произвольная аффинная система координат (п. VII.9), которые мы будем обозначать x_1, x_2, с началом координат O и базисными векторами $\vec{p}_1$, $\vec{p}_2$, так что радиус-вектор $\vec{r}=x_1\vec{p}_1+x_2\vec{p}_2$. В плоскости $(\overline{P})$ также произвольно выбрана аффинная система координат y_1, y_2, с началом координат $\overline{O}$ и базисными векторами $\vec{q}_1$, $\vec{q}_2$. Обозначим координаты точки $f(O)$ в плоскости $(\overline{P})$ через b_1, b_2. Пусть даны координаты x_1, x_2 некоторой точки M плоскости (P); каковы будут координаты y_1, y_2 соответствующей точки $f(M)$? Так как

$$\overrightarrow{\overline{O}f(M)}=\overrightarrow{\overline{O}f(O)}+\overrightarrow{f(O)f(M)}=b_1\vec{q}_1+b_2\vec{q}_2+A(\overrightarrow{OM}),$$

то в силу выведенных выше формул для преобразования координат вектора получаем

$$\left.\begin{aligned} y_1&=a_{11}x_1+a_{12}x_2+b_1,\\ y_2&=a_{21}x_1+a_{22}x_2+b_2, \end{aligned}\right\} \tag{25}$$

коротко,

$$\mathbf{y}=\mathbf{A}\mathbf{x}+\mathbf{b}, \tag{26}$$

где $\mathbf{A}=\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$ — матрица отображения А в выбранных базисах, а $\mathbf{b}=\begin{pmatrix} b_1\\ b_2\end{pmatrix}$. Легко проверить, что и обратно, если координаты точек преобразуются по формулам (25), то преобразование обладает свойствами, описанными в предыдущем абзаце. Наиболее простые формулы получаются, если при отображении начало координат плоскости (P) переходит в начало координат плоскости $(\overline{P})$: тогда $b_1=b_2=0$, т. е. формулы (25) принимают вид

$$\left.\begin{aligned} y_1&=a_{11}x_1+a_{12}x_2,\\ y_2&=a_{21}x_1+a_{22}x_2, \end{aligned}\right\} \tag{27}$$

коротко, $\mathbf{y}=\mathbf{A}\mathbf{x}$. По этим же формулам преобразуются в общем случае (25) координаты векторов.

Если вместо плоскости рассматривается пространство или числовое пространство любой размерности (п. VII.18), то вместо (25) получаются аналогичные формулы, но с другим числом строк и столбцов. Краткая запись имеет тот же вид (26), где матрица $\mathbf{A}$, вообще говоря, прямоугольная, так как отображаться друг в друга могут пространства различных размерностей.

Пусть всё же эти размерности одинаковы, для простоты рассмотрим опять отображение плоскости (P) в плоскость $(\overline{P})$. Если $\det \mathbf{A} \neq 0$, то отображение называется *аффинным*. В этом случае, умножив равенство (26) слева на $\mathbf{A}^{-1}$, получаем $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} - \mathbf{A}^{-1}\mathbf{b}$, т. е. равенство того же вида, что (26). Значит, *обратное отображение* плоскости $(\overline{P})$ в плоскость (P) также является аффинным.

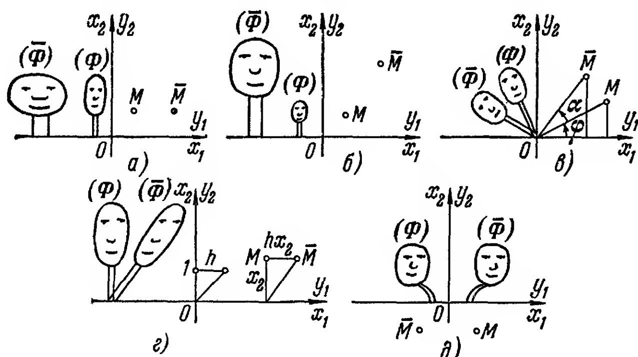


Рис. 223.

Аффинные отображения плоскости:

а) растяжение вдоль оси x_1 в k раз

$$y_1 = kx_1, \quad y_2 = x_2 \qquad \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$$

б) всестороннее растяжение в k раз

$$y_1 = kx_1, \quad y_2 = kx_2 \qquad \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$$

в) поворот на угол α

$$y_1 = x_1 \cos\alpha - x_2 \sin\alpha, \quad y_2 = x_1 \sin\alpha + x_2 \cos\alpha \qquad \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

г) сдвиг вдоль оси x_1

$$y_1 = x_1 + hx_2, \quad y_2 = x_2 \qquad \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}$$

д) зеркальное отражение от оси x_2

$$y_1 = -x_1, \quad y_2 = x_2 \qquad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Наиболее распространенные примеры аффинных отображений плоскости на себя, при которых начало координат остается на месте, показаны на рис. 223. Там же указаны формулы преобразования координат и соответствующие матрицы в декартовой системе координат. (Докажите формулы для третьего примера, исходя из того, что $y_1 = \rho\cos(\varphi + \alpha)$, $y_2 = \rho\sin(\varphi + \alpha)$, где $\rho = OM = O\overline{M}$.) Конечно, возможны и комбинации этих простейших отображений, а также дополнительные параллельные переносы.

Если $\det \mathbf{A} = 0$, то ранг матрицы $\mathbf{A}$ равен 1 или 0; в первом случае, как мы видели выше, плоскость (P) отображается на прямую (так будет, в частности, при проектировании), а во втором — на точку плоскости $(\overline{P})$.

Аналогично п. II.7 легко доказать, что при аффинном отображении плоскости порядок алгебраической линии не меняется. В частности, прямые линии, как единственные линии первого порядка (п. II.9), при аффинном отображении остаются прямыми. Так как точка пересечения двух линий при

отображении должна переходить в точку пересечения их образов, то пересекающиеся прямые переходят в пересекающиеся, а следовательно, параллельные — в параллельные. Легко проверить из определения, что отношение параллельных отрезков при аффинном отображении сохраняется; однако отношение непараллельных отрезков, длины и углы, вообще говоря, изменяются. Линии второго порядка при аффинном отображении переходят также в линии второго порядка. При этом эллипс, как единственная целиком конечная линия второго порядка, переходит в эллипс (или, в частности, в окружность); парабола, как бесконечная линия второго порядка, состоящая из одного куска, переходит в параболу, гипербола переходит в гиперболу.

Вернемся к линейным отображениям линейных пространств (см. начало этого пункта). Пусть даны два таких отображения, A и B пространства (R) в (S). Тогда можно определить *сумму* этих *отображений* по формуле $(\mathrm{A}+\mathrm{B})\vec{x}=\mathrm{A}\vec{x}+\mathrm{B}\vec{x}$; аналогично определяется *умножение отображения на число*: $(\lambda\mathrm{A})\vec{x}=\lambda(\mathrm{A}\vec{x})$. Легко проверить, что если в (R) и (S) выбраны базисы, то при указанных действиях те же действия производятся над матрицами этих отображений, причем выполняются все аксиомы линейных действий; роль нуля играет отображение всего (R) в нуль-вектор пространства (S).

Умножение отображений по определению означает их последовательное выполнение. Более точно, пусть дано отображение B пространства (R) в (S) и отображение A пространства (S) в (T). Тогда под AB понимается «сложное» отображение (R) в (T), получающееся в результате выполнения сначала отображения B, а затем отображения A, т. е. $(\mathrm{AB})\vec{x}=\mathrm{A}(\mathrm{B}\vec{x})$. Если в (R), (S) и (T) выбраны базисы, то при умножении преобразований их матрицы также перемножаются; *отсюда и проистекает данное в п. 2 правило умножения матриц*. Из-за важности этого правила дадим его вывод, хотя бы на примере квадратных матриц второго порядка. Пусть отображения B и A задаются соответственно формулами

$$\left.\begin{aligned} y_1&=b_{11}x_1+b_{12}x_2,\\ y_2&=b_{21}x_1+b_{22}x_2,\end{aligned}\right\}\qquad \left.\begin{aligned} z_1&=a_{11}y_1+a_{12}y_2,\\ z_2&=a_{21}y_1+a_{22}y_2;\end{aligned}\right\}$$

тогда для получения «сложного» отображения надо первые формулы подставить во вторые и мы получим

$$z_1=a_{11}(b_{11}x_1+b_{12}x_2)+a_{12}(b_{21}x_1+b_{22}x_2)=$$
$$=(a_{11}b_{11}+a_{12}b_{21})x_1+(a_{11}b_{12}+a_{12}b_{22})x_2,$$
$$z_2=(\text{аналогично})=(a_{21}b_{11}+a_{22}b_{21})x_1+(a_{21}b_{12}+a_{22}b_{22})x_2,$$

т. е. мы приходим к матрице AB, определенной по правилу п. 2.

Из связи умножения матриц с умножением преобразований можно было бы вывести указанные в п. 2 свойства действий с матрицами, так как соответствующие свойства преобразований легко выводятся из их линейности, Отметим только свойство $\mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C}$, вытекающее из соответствующего свойства отображений $\mathrm{A}(\mathrm{BC})=(\mathrm{AB})\mathrm{C}$: как в левой, так и в правой части последнего равенства стоит отображение, которое получится, если выполнить сначала отображение C, затем B и, наконец, A. При отображениях пространства в себя, т. е. при $(S)=(R)$ роль единицы при умножении отображений играет *тождественное отображение* I, при котором каждый вектор переходит в себя $\mathrm{I}\vec{x}=\vec{x}$: ясно, что всегда $\mathrm{AI}=\mathrm{IA}=\mathrm{A}$. Этому отображению при любом выборе базиса отвечает единичная матрица **I**. Обратная матрица отвечает обратному отображению. Нетрудно понять, почему умножение матриц оказалось, вообще говоря, неперестановочным: ведь одни и те же отображения приходится выполнять в разном порядке, что может существенно изменить результат. (Проверьте, например, что если к точке (0; 1) применить

отображение рис. 223 *а*) с $k=2$, а затем отображение *в*) с $\alpha=90°$, то получится точка (2; 0), если же эти действия выполнить в обратном порядке, получится точка (1; 0).)

Вообще, какие-либо два действия могут либо *коммутировать*, если результат их совместного выполнения не зависит от порядка, в котором они выполняются, либо не коммутировать. (Коммутируют ли следующие действия: а) наполнение бассейна водой; б) прыжок в этот бассейн с вышки?)

7. Преобразование матрицы отображения при замене базиса. Мы уже отмечали, что при рассмотрении линейного отображения A пространства (R) в пространство (S) матрица этого отображения зависит от выбора базисов в обоих пространствах. Одно и то же отображение может в одном базисе иметь более сложную матрицу, а в другом — более простую. Сейчас мы установим, какова эта зависимость, вернувшись для определенности к выводу формулы (24). Пусть в (R) выбран новый базис $\vec{p}_1'$, $\vec{p}_2'$, $\vec{p}_3'$. Тогда любой вектор $\vec{x}$ можно разложить как по старому, так и по новому базису:

$$\vec{x}=x_1\vec{p}_1+x_2\vec{p}_2+x_3\vec{p}_3=x_1'\vec{p}_1'+x_2'\vec{p}_2'+x_3'\vec{p}_3', \tag{28}$$

где x_1, x_2, x_3 — старые, а x_1', x_2', x_3' — новые координаты вектора $\vec{x}$. Каждый из новых базисных векторов можно разложить по старым:

$$\left.\begin{aligned}\vec{p}_1'&=h_{11}\vec{p}_1+h_{21}\vec{p}_2+h_{31}\vec{p}_3,\\ \vec{p}_2'&=h_{12}\vec{p}_1+h_{22}\vec{p}_2+h_{32}\vec{p}_3,\\ \vec{p}_3'&=h_{13}\vec{p}_1+h_{23}\vec{p}_2+h_{33}\vec{p}_3,\end{aligned}\right\} \tag{29}$$

где h_{ij} — некоторые коэффициенты, определяющие переход от старого базиса к новому. Если подставить формулы (29) в (28), а затем приравнять коэффициенты при одинаковых базисных векторах $\vec{p}_1$, $\vec{p}_2$, $\vec{p}_3$, мы получим (проверьте!)

$$\left.\begin{aligned}x_1&=h_{11}x_1'+h_{12}x_2'+h_{13}x_3',\\ x_2&=h_{21}x_1'+h_{22}x_2'+h_{23}x_3',\\ x_3&=h_{31}x_1'+h_{32}x_2'+h_{33}x_3'.\end{aligned}\right\} \tag{30}$$

Отметим, что *матрица*

$$\mathbf{H}=\begin{pmatrix}h_{11} & h_{12} & h_{13}\\ h_{21} & h_{22} & h_{23}\\ h_{31} & h_{32} & h_{33}\end{pmatrix}$$

перехода от новых координат к старым обязательно невырожденная, так как, задавая x_1, x_2, x_3, мы должны получить вполне определенные значения x_1', x_2', x_3'.

Формулы (30), наподобие формул (23), можно записать кратко в виде

$$\mathbf{x} = \mathbf{H}\mathbf{x}', \quad \text{где } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{x}' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix}.$$

Подобным образом, если в пространстве (S) введен новый базис $\vec{q}_1'$, $\vec{q}_2'$ с матрицей перехода $\mathbf{K}$ от старых координат к новым, то $\mathbf{y} = \mathbf{K}\mathbf{y}'$. Подставляя эти формулы в (24), получим $\mathbf{K}\mathbf{y}' = \mathbf{A}\mathbf{H}\mathbf{x}'$, откуда, умножая слева на $\mathbf{K}^{-1}$, получим

$$\mathbf{y}' = \mathbf{K}^{-1}\mathbf{A}\mathbf{H}\mathbf{x}', \text{ т. е. } \mathbf{y}' = \mathbf{A}'\mathbf{x}', \text{ где } \mathbf{A}' = \mathbf{K}'\mathbf{A}\mathbf{H}.$$

Это и есть матрица рассматриваемого отображения в новых базисах.

В частности, если $(S) = (R)$, то $\mathbf{K} = \mathbf{H}$ и потому

$$\mathbf{A}' = \mathbf{H}^{-1}\mathbf{A}\mathbf{H}. \tag{31}$$

Сейчас мы выясним, каков *геометрический смысл определителя матрицы аффинного отображения плоскости на себя.* Пусть отображение определяется формулами (25), причем допустим сначала, что базис $\vec{p}_1$, $\vec{p}_2$, выбранный в плоскости (P), является декартовым; по условию $(\overline{P}) = (P)$. Так как координаты векторов преобразуются по формулам (27), то векторы $\vec{p}_1$ и $\vec{p}_2$ перейдут соответственно в векторы $\vec{s}_1 = a_{11}\vec{p}_1 + a_{21}\vec{p}_2$ и $\vec{s}_2 = a_{12}\vec{p}_1 + a_{22}\vec{p}_2$ (почему?). Площадь квадрата, построенного на векторах $\vec{p}_1$, $\vec{p}_2$, равняется единице, а площадь параллелограмма, в который отображается этот квадрат, т. е. построенного на векторах $\vec{s}_1$, $\vec{s}_2$, в силу конца п. VII.13, равна $\left| \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} \right| = |\det \mathbf{A}|$. Если теперь заметить, что при аффинном отображении все участки плоскости преобразуются одинаково и потому площади любых фигур меняются в одной и той же пропорции, то мы получаем, что $|\det \mathbf{A}|$ как раз равен коэффициенту изменения площадей при данном отображении Знак определителя $\det \mathbf{A}$ также имеет геометрический смысл. Если он положительный, то направление обхода контура фигуры при отображении сохраняется, т е. если мы обходим контур прообраза в положительном направлении, то и контур образа проходится в положительном направлении. Если же $\det \mathbf{A} < 0$, то при отображении направление обхода контура меняется на противоположное. (Проверьте все эти утверждения на примерах рис. 223)

Если теперь перейти к произвольному базису, то

$$\det \mathbf{A}' = \det (\mathbf{H}^{-1}\mathbf{A}\mathbf{H}) = \det (\mathbf{H}^{-1}) \det \mathbf{A} \det \mathbf{H} = (\det \mathbf{H})^{-1} \det \mathbf{A} \det \mathbf{H} = \det \mathbf{A}.$$

Мы видим, что хотя матрица аффинного отображения зависит от выбора базиса, но ее определитель от этого не зависит, т. е. его смысл одинаков при любом выборе базиса

Если рассматривается аффинное отображение одной плоскости на другую, то $|\det \mathbf{A}|$ также равняется коэффициенту изменения площадей, если площадь в каждой плоскости измерять через площадь параллелограмма, построенного на базисных векторах.

Если рассматривается аффинное отображение пространства на себя, то аналогично можно показать, что $|\det \mathbf{A}|$ равняется коэффициенту изменения объемов, а знак $\det \mathbf{A}$ имеет $+$ или $-$ в зависимости от того, остаются при этом отображении правые тройки векторов правыми или переходят в левые.

Подобным образом можно истолковать смысл определителя матрицы аффинного отображения числового пространства любой размерности на себя. Из этого смысла, в частности, сразу вытекает формула (5), так как ясно, что при последовательном выполнении аффинных преобразований соответствующие коэффициенты изменения объемов перемножаются.

8. Матрица отображения в базисе из собственных векторов. Рассмотрим линейное отображение A линейного пространства (R) в себя. Если некоторый ненулевой вектор $\vec{x}$ преобразуется в параллельный вектор, так что для $\vec{x}$ преобразование состоит в умножении на скаляр λ, т. е. $A\vec{x}=\lambda\vec{x}$, то $\vec{x}$ называется *собственным вектором отображения* A, *отвечающим собственному значению* λ.

Так, для первого примера на рис. 223 собственным является любой вектор, параллельный оси x_1 (он отвечает собственному значению k) или оси x_2 (он не меняется, т. е. отвечает собственному значению 1). Мы предоставим читателю найти собственные векторы и собственные значения для других примеров на рис. 223.

Если в (R) выбран базис, то вместо вектора $\vec{x}$ можно рассматривать числовой вектор $\mathbf{x}$ из его координат, и в силу п. 6 равенство, определяющее собственный вектор, приобретает вид $\mathbf{Ax}=\lambda\mathbf{x}$, т. е. (11). Значит, в силу п. 4 вектор $\mathbf{x}$ должен быть собственным вектором матрицы $\mathbf{A}$ рассматриваемого отображения в выбранном базисе. В п. 4 мы разобрали, как находить эти векторы. На основании п. 4 мы заключаем, что преобразование A имеет столько собственных значений, какова размерность пространства (R), однако среди этих значений могут быть мнимые, а также совпадающие. Так, в примере *в*) рис. 223 собственные значения — мнимые (проверьте!), поэтому ни один ненулевой вектор не сохраняет при отображении параллельности. (Впрочем, из п. VIII.8 легко вывести, что если размерность (R) нечетная, то уравнение (13) имеет по крайней мере одно вещественное решение, а потому имеется по крайней мере один собственный вектор.)

Допустим для определенности, что пространство (R) трехмерно, и предположим, что у отображения A можно указать три линейно независимых (вещественных) собственных вектора $\vec{l}_1$, $\vec{l}_2$, $\vec{l}_3$. Примем эти векторы за базис; оказывается, что тогда матрица отображения $\mathbf{A}$ приобретает особенно простой вид. В самом деле, пусть мы еще не знаем, какая она, и напишем пока

$$\left.\begin{aligned} y'_1 &= a'_{11}x'_1 + a'_{12}x'_2 + a'_{13}x'_3, \\ y'_2 &= a'_{21}x'_1 + a'_{22}x'_2 + a'_{23}x'_3, \\ y'_3 &= a'_{31}x'_1 + a'_{32}x'_2 + a'_{33}x'_3, \end{aligned}\right\} \qquad (32)$$

где штрихом обозначены координаты в указанном базисе. Так как вектор $\vec{l}_1$ имеет координаты $x'_1=1$, $x'_2=0$, $x'_3=0$, а после отображения он переходит в вектор $\lambda_1\vec{l}_1$ с координатами $y'_1=\lambda_1$, $y'_2=0$,

$y'_3=0$, то должно быть

$$\lambda_1=a'_{11}1+a'_{12}0+a'_{13}0,$$
$$0=a'_{21}1+a'_{22}0+a'_{23}0,$$
$$0=a'_{31}1+a'_{32}0+a'_{33}0,$$

откуда находим, что $a'_{11}=\lambda_1$, $a'_{21}=0$, $a'_{31}=0$. Аналогично получаем (проверьте!), что $a'_{22}=\lambda_2$, $a'_{33}=\lambda_3$, $a'_{12}=a'_{13}=a'_{23}=a'_{32}=0$, т. е. формулы (32) на самом деле имеют вид

$$y'_1=\lambda_1x'_1,\quad y'_2=\lambda_2x'_2,\quad y'_3=\lambda_3x'_3.$$

Стало быть, *матрица линейного отображения в базисе, состоящем из собственных векторов, имеет диагональный вид*

$$\mathbf{A}'=\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix}=\operatorname{diag}(\lambda_1,\ \lambda_2,\ \lambda_3).$$

В силу п. 4 для наличия такого базиса достаточно, чтобы все корни характеристического уравнения матрицы **A** были вещественными и различными. Так как любую матрицу можно рассматривать как матрицу некоторого линейного отображения, а изменение матрицы такого отображения при замене базиса осуществляется по формуле (31), то полученный результат можно сформулировать так: *если у квадратной матрицы* **A** *все корни характеристического уравнения вещественны и различны, то можно подобрать такую невырожденную матрицу* **H**, *для которой матрица* $\mathbf{H}^{-1}\mathbf{A}\mathbf{H}$ *будет диагональной, причем на диагонали будут стоять как раз указанные корни.*

Если у характеристического уравнения не все корни вещественные и для некоторого мнимого корня найден соответствующий собственный вектор из уравнений (12) (координаты этого вектора также получатся мнимыми), то такой собственный вектор не будет иметь непосредственного геометрического смысла. Так будет, например, для третьего отображения на рис. 223. Однако можно пользоваться комплексными числами, не обязательно давая им непосредственное геометрическое истолкование. Если допускать, что рассматриваемые проекции могут принимать и комплексные значения, то все приведенные выкладки остаются в силе. В частности, утверждение предыдущего абзаца остается справедливым для любой квадратной матрицы, у которой все собственные значения (корни характеристического уравнения) различные; однако в общем случае матрица **H** может получиться комплексной.

Если у характеристического уравнения заданной матрицы имеются равные корни, то в общем случае такая матрица к диагональному виду не приводится. Так будет, в частности, в четвертом примере рис. 223.

Отметим в заключение, что при замене базиса матрица **A** меняется по формуле (31), но характеристическое уравнение (13) остается при этом неизменным. В самом деле,

$$\det(\mathbf{A}'-\lambda\mathbf{I})=\det(\mathbf{H}^{-1}\mathbf{A}\mathbf{H}-\lambda\mathbf{I})=\det[\mathbf{H}^{-1}(\mathbf{A}-\lambda\mathbf{I})\,\mathbf{H}]=$$
$$=\det(\mathbf{H}^{-1})\det(\mathbf{A}-\lambda\mathbf{I})\det\mathbf{H}=\frac{1}{\det\mathbf{H}}\det(\mathbf{A}-\lambda\mathbf{I})\det\mathbf{H}=\det(\mathbf{A}-\lambda\mathbf{I}).\qquad(33)$$

9. Замена декартова базиса. В этом пункте мы будем считать, что (R) является не только линейным, но евклидовым пространством (пп. VII. 20—21), так что можно говорить о декартовых базисах в (R). Выясним, какова матрица перехода от одного декартова базиса

к другому, причем для определенности будем считать (R) трехмерным. Тогда надо пользоваться формулами (29). Пусть $\vec{p}_1$, $\vec{p}_2$, $\vec{p}_3$ образуют декартов базис, т. е. играют ту же роль, что векторы **i**, **j**, **k** в § VII.3. Для того чтобы $\vec{p}'_1$, $\vec{p}'_2$, $\vec{p}'_3$ также образовывали декартов базис, необходимо и достаточно, чтобы

$$\vec{p}'_1\cdot\vec{p}'_1=\vec{p}'_2\cdot\vec{p}'_2=\vec{p}'_3\cdot\vec{p}'_3=1,\quad \vec{p}'_1\cdot\vec{p}'_2=\vec{p}'_2\cdot\vec{p}'_3=\vec{p}'_1\cdot\vec{p}'_3=0$$

(почему?). Раскрывая эти скалярные произведения по формуле (VII.12), получим

$$h_{11}^2+h_{21}^2+h_{31}^2=h_{12}^2+h_{22}^2+h_{32}^2=h_{13}^2+h_{23}^2+h_{33}^2=1,$$
$$h_{11}h_{12}+h_{21}h_{22}+h_{31}h_{32}=h_{12}h_{13}+h_{22}h_{23}+h_{32}h_{33}=$$
$$=h_{11}h_{13}+h_{21}h_{23}+h_{31}h_{33}=0.$$

Совокупность этих шести равенств можно сокращенно записать в виде

$$\begin{pmatrix}h_{11} & h_{21} & h_{31}\\ h_{12} & h_{22} & h_{32}\\ h_{13} & h_{23} & h_{33}\end{pmatrix}\cdot\begin{pmatrix}h_{11} & h_{12} & h_{13}\\ h_{21} & h_{22} & h_{23}\\ h_{31} & h_{32} & h_{33}\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}.$$

(Проверьте это, перемножив в левой части матрицы по правилам п. 2 и сравнив результат с правой частью.) С помощью обозначений пп. 1 и 3 это условие можно записать в виде

$$\mathbf{H}^*\mathbf{H}=\mathbf{I},\ \text{т. е.}\ \mathbf{H}^*=\mathbf{H}^{-1}. \qquad (34)$$

Матрица, удовлетворяющая этим равенствам, т. е. обратная к своей транспонированной, называется *ортогональной*. Доказанное нами свойство можно сформулировать так: *матрица перехода от одной декартовой системы к другой является ортогональной; обратно, если матрица перехода ортогональная, то из декартова базиса обязательно получается декартов базис*. (Проверьте, что матрицы переходов (II.3) и (II.4) ортогональные.)

Ортогональные матрицы появляются также в связи с отображениями. Именно, линейное отображение евклидова пространства в себя называется *ортогональным*, если при этом отображении длины векторов не изменяются. Так как при таком отображении любой треугольник переходит в равный треугольник по так называемому «третьему признаку равенства треугольников», то и все углы при этом сохраняются Ортогональное отображение может быть либо движением всего пространства как целого, либо комбинацией движения и зеркального отражения.

Например, из отображений плоскости, указанных на рис. 223, ортогональными являются третье (движение) и пятое (отражение).

Исходя из формул, аналогичных (22), и рассуждая, как в начале этого пункта, легко проверить, что матрица **A** ортогонального преобразования в декартовом базисе ортогональная, и обратно, если в декартовом базисе преобразование имеет ортогональную матрицу, то оно ортогональное.

Из равенства $\mathbf{A}^*\mathbf{A}=\mathbf{I}$ (см. (34)) вытекает $\det A^*\cdot\det\mathbf{A}=(\det\mathbf{A})^2=\det\mathbf{I}=1$, откуда $\det A=\pm 1$. Это вытекает и из геометрического смысла определителя матрицы линейного отображения (п. 7). При этом для движения определи-

тель равен 1, а для зеркального отражения или для комбинации движения и отражения он равен -1.

Отметим одно следствие, которое применяется в механике. Пусть задано движение (обычного) пространства, при котором начало координат остается на месте. Это движение можно рассматривать как ортогональное преобразование A совокупности всех векторов трехмерного пространства. Если левую часть характеристического уравнения разложить по формуле (VIII.25),

$$\det(\mathbf{A}-\lambda\mathbf{I}) = -(\lambda-\lambda_1)(\lambda-\lambda_2)(\lambda-\lambda_3),$$

а затем в этом тождестве положить $\lambda=0$, мы получим, что $\lambda_1\lambda_2\lambda_3=1$. Отсюда следует, что по крайней мере одно собственное значение λ_k будет вещественным положительным, т. е. существует вектор $\vec{x}_0 \neq \vec{0}$, для которого $A\vec{x}_0=\lambda_k\vec{x}_0$; так как длины сохраняются, то $\lambda_k=1$. Значит, рассматриваемое движение представляет собой поворот пространства вокруг оси, проходящей через начало координат параллельно собственному вектору, отвечающему собственному значению, равному 1.

10. Симметрические матрицы. Укажем применение полученных результатов к исследованию симметрических матриц (см. п. 1).

Можно доказать следующие свойства симметрических матриц.

У симметрической матрицы все собственные значения вещественные.

Например, симметрическая матрица второго порядка имеет вид

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}; \tag{35}$$

ее характеристическое уравнение

$$\begin{vmatrix} a-\lambda & b \\ b & c-\lambda \end{vmatrix} = 0,$$

или

$$\lambda^2-(a+c)\lambda+ac-b^2=0 \tag{36}$$

имеет корни

$$\lambda_{1,2}=\frac{a+c}{2}\pm\sqrt{\frac{(a+c)^2}{4}-ac+b^2}=\frac{a+c}{2}\pm\sqrt{\frac{(a-c)^2}{4}+b^2}, \tag{37}$$

очевидно, вещественные. Доказательства этого утверждения, как и двух дальнейших, для матриц выше второго порядка мы не приводим.

Собственные векторы, отвечающие различным собственным значениям, обязательно ортогональны друг другу.

Рассмотрим, например, матрицу (35) при $b \neq 0$. Координаты собственного вектора определяются из системы (12), которая в нашем случае имеет вид

$$\left.\begin{aligned} (a-\lambda)x_1+bx_2&=0,\\ bx_1+(c-\lambda)x_2&=0. \end{aligned}\right\}$$

Если λ — собственное значение, то эти два уравнения зависимы (п. VI.6) и потому достаточно ограничиться одним из них, например первым. Чтобы удовлетворить ему, можно положить $x_1=-b$, $x_2=a-\lambda$. Полагая $\lambda=\lambda_1$, $\lambda=\lambda_2$, мы получим два собственных вектора

$$\begin{pmatrix} -b \\ a-\lambda_1 \end{pmatrix} \text{ и } \begin{pmatrix} -b \\ a-\lambda_2 \end{pmatrix}. \tag{38}$$

Их скалярное произведение равно

$$b^2+(a-\lambda_1)(a-\lambda_2)=\lambda_1\lambda_2-a(\lambda_1+\lambda_2)+b^2+a^2=ac-b^2-a(a+c)+b^2+a^2=0$$

(при выводе мы воспользовались формулами для суммы и произведения корней квадратного уравнения (36)), откуда и вытекает перпендикулярность векторов (38). Если же $b=0$, то собственными векторами служат $\begin{pmatrix}1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix}$ (продумайте это!), так что и в этом случае они перпендикулярны

Если не все корни характеристического уравнения различны, т. е. некоторое собственное значение повторяется k раз, то можно указать k взаимно ортогональных собственных векторов, отвечающих этому собственному значению.

Например, для матрицы (35) из формулы (37) видно, что собственное значение будет двойным, только если $a=c$ и $b=0$. Но тогда все векторы будут собственными (проверьте!), так что, конечно, среди них можно выбрать два взаимно перпендикулярных.

Из указанных свойств вытекает, что если заданную симметрическую матрицу $\mathbf{A}$ рассматривать как матрицу линейного преобразования в некотором декартовом базисе, то всегда можно указать новый декартов базис, составленный целиком из собственных векторов матрицы $\mathbf{A}$. Например, для трехмерного случая характеристическое уравнение будет третьей степени и поэтому имеет три корня, которые, как было сказано выше, все вещественные. Если они все различные, то соответствующие им собственные векторы взаимно перпендикулярны, и потому если их взять единичной длины, то и получится искомый базис. Если $\lambda_1=\lambda_2\neq\lambda_3$, то можно выбрать два перпендикулярных собственных вектора, отвечающих значению λ_1, а вектор, отвечающий значению λ_3, будет перпендикулярен к ним обоим. Наконец, если все три собственных значения одинаковы, то для этого общего значения можно указать три взаимно перпендикулярных собственных вектора.

Так как переход от одного декартова базиса к другому осуществляется с помощью ортогональной матрицы (п. 9), а при базисе из собственных векторов матрица после преобразования по формуле (31) становится диагональной (см. п. 8), то доказанное в предыдущем абзаце на языке матриц можно сформулировать так: *для каждой симметрической матрицы* $\mathbf{A}$ *можно подобрать такую ортогональную матрицу* $\mathbf{H}$, *для которой матрица* $\mathbf{H}^{-1}\mathbf{A}\mathbf{H}$ *будет диагональной, причем на диагонали будут стоять собственные значения матрицы* $\mathbf{A}$.

§ 3. Квадратичные формы

11. Квадратичные формы. Квадратичной формой от нескольких переменных называется однородный многочлен второй степени от этих переменных. Например, квадратичная форма от трех переменных

x_1, x_2, x_3 в общем случае имеет вид

$$F = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3, \qquad (39)$$

где a_{11}, a_{22}, ..., a_{23} — некоторые числовые коэффициенты (а двойки поставлены для упрощения получающихся формул). *Матрицей* такой формы называется симметрическая матрица

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}.$$

С помощью этой матрицы форму (39) можно переписать в виде

$$F = (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)\,x_1 + (a_{12}x_1 + a_{22}x_2 + a_{23}x_3)\,x_2 +$$
$$+ (a_{13}x_1 + a_{23}x_2 + a_{33}x_3)\,x_3 = y_1x_1 + y_2x_2 + y_3x_3 = (x_1 \;\; x_2 \;\; x_3)\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

где

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{12}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{13}x_1 + a_{23}x_2 + a_{33}x_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{A}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Значит, если ввести числовой вектор $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, то получим

$$F = \mathbf{x}^*\mathbf{A}\mathbf{x}. \qquad (40)$$

Обратно, если форма представлена в виде (40), причем матрица $\mathbf{A}$ симметрическая, то $\mathbf{A}$ является матрицей этой формы.

Проведем теперь какую угодно линейную замену переменных вида (30) или, коротко,

$$\mathbf{x} = \mathbf{H}\mathbf{x}'. \qquad (41)$$

Тогда в силу формулы (4) $\mathbf{x}^* = \mathbf{x}'^*\mathbf{H}^*$, откуда

$$F = \mathbf{x}'^*\mathbf{H}^*\mathbf{A}\mathbf{H}\mathbf{x}' = \mathbf{x}'^*(\mathbf{H}^*\mathbf{A}\mathbf{H})\,\mathbf{x}',$$

т. е.

$$F = \mathbf{x}'^*\mathbf{A}'\mathbf{x}', \text{ где } \mathbf{A}' = \mathbf{H}^*\mathbf{A}\mathbf{H}. \qquad (42)$$

Однако матрица $\mathbf{A}'$ симметрическая, так как в силу формулы (4)

$$\mathbf{A}'^* = (\mathbf{H}^*\mathbf{A}\mathbf{H})^* = \mathbf{H}^*\mathbf{A}^*\mathbf{H}^{**} = \mathbf{H}^*\mathbf{A}\mathbf{H} = \mathbf{A}'.$$

Значит, $\mathbf{A}'$ и служит матрицей квадратичной формы после замены переменных.

Итак, при подстановке (41) матрица квадратичной формы преобразуется по формуле (42). Если, в частности, считать матрицу $\mathbf{H}$ ортогональной, то в силу формулы (34) получаем, что $\mathbf{A}' = \mathbf{H}^{-1}\mathbf{A}\mathbf{H}$. Как было показано в п. 10, всегда можно подобрать матрицу $\mathbf{H}$

так, чтобы $\mathbf{A}' = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$, где на диагонали стоят собственные значения матрицы **A**. Но тогда квадратичная форма F в новых переменных приобретает *диагональный вид*

$$F = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \lambda_2 x_3'^2. \tag{43}$$

Итак, *квадратичную форму* (39) *можно с помощью преобразования* (30) *с ортогональной матрицей привести к диагональному виду* (43), *где* λ_1, λ_2, λ_3 — *собственные значения матрицы* **A**.

Указанное формальное преобразование имеет следующий геометрический смысл. Рассмотрим матрицу **A** как матрицу некоторого линейного преобразования A в декартовом базисе с координатами x_1, x_2, x_3. Тогда переход (41), приводящий квадратичную форму F к виду (43), отвечает переходу к новому базису, составленному из собственных векторов преобразования A.

В п. 8 (формула (33)) мы показали, что при любой замене вида (31) определитель $\det(\mathbf{A} - \lambda \mathbf{I})$ не меняется. Значит, если разложить этот определитель по степеням λ, то коэффициенты при этих степенях также не будут меняться; эти коэффициенты будут инвариантными относительно замены декартовых координат. Например, квадратичная форма от двух переменных при обозначениях, принятых в аналитической геометрии, имеет вид

$$Ax^2 + 2Bxy + Cy^2,$$

т. е. ее матрица такова:

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix},$$

а характеристическое уравнение

$$\begin{vmatrix} A-\lambda & B \\ B & C-\lambda \end{vmatrix} = \lambda^2 - (A+C)\lambda + AC - B^2.$$

Значит, выражения $A+C$ и $AC-B^2$ инвариантны относительно замены декартовых координат (см. п. II.13).

12. Упрощение уравнений линий и поверхностей второго порядка. Преобразование квадратичной формы, описанное в п. 11, применяется, в частности, к упрощению уравнений линий и поверхностей второго порядка. Остановимся на уравнении поверхностей, так как упрощение уравнений линий второго порядка описано в п. II.13.

Пусть поверхность второго порядка задана в обычном для аналитической геометрии виде (X.13). Переход к новой декартовой системе координат с тем же началом, как было показано в п. 9, сводится к замене переменных

$$\left.\begin{aligned} x &= h_{11}x' + h_{12}y' + h_{13}z', \\ y &= h_{21}x' + h_{22}y' + h_{23}z', \\ z &= h_{31}x' + h_{32}y' + h_{33}z' \end{aligned}\right\} \tag{44}$$

с ортогональной матрицей перехода **H**. (Ясно, что если начало координат не заменяется, то координаты точек преобразуются по тем же формулам, что координаты векторов.) При подстановке этих выражений в уравнение (X.13) группа членов второй степени и группа членов первой степени преобразуются независимо друг от друга. Если следить сначала только за группой членов второй степени (квадратичной формой), то на основании п. 9, мы получаем, что всегда можно выбрать систему координат x', y', z' так, что эта группа членов приобретет диагональный вид. Поэтому все уравнение после преобразования будет иметь вид

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 + G'x' + H'y' + I'z' + J = 0; \qquad (45)$$

здесь λ_1, λ_2, λ_3 — корни уравнения

$$\begin{vmatrix} A-\lambda & B & D \\ B & C-\lambda & E \\ D & E & F-\lambda \end{vmatrix} = 0,$$

а G', H', I' — некоторые новые коэффициенты при членах первой степени, которые сами получатся после подстановки (44). Уравнение (45) — это уравнение (X.14) в других обозначениях; его исследование описано в п. X.11.

§ 4. Нелинейные отображения

13. Общие понятия. Начнем с отображения плоскости в плоскость. Для этого предположим, что имеются две плоскости (P) и $(\overline{P})$, которые, впрочем, могут быть совмещенными. Пусть каждой точке M из плоскости (P) (или из некоторой ее области) по определенному закону соответствует определенная точка $\overline{M}$ плоскости $(\overline{P})$. Тогда говорят, что задано *отображение плоскости* (P) (или ее области) *в плоскость* $(\overline{P})$. Как было отмечено в конце п. VIII.11 при рассмотрении отображений специального вида, при отображении (P) в $(\overline{P})$ линии переходят в линии; фигуры в плоскости (P) переходят в фигуры в плоскости $(\overline{P})$, хотя при этом переходе форма фигуры может так измениться (рис. 224), что по *прообразу* в плоскости (P) будет трудно узнать *образ* в плоскости $(\overline{P})$ и наоборот. Возможны также случаи *вырождения*, когда, например, отдельные фигуры при отображении «сплющиваются» в линии или даже в точки.

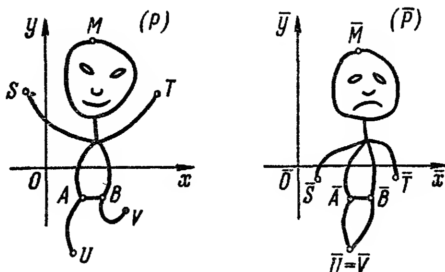


Рис. 224.

Иногда приходится рассматривать *обратное отображение*, которое получится, если задавать образ в $(\overline{P})$ и искать прообраз в (P). При этом, подобно обратной функции (п. I.21), может получиться осложнение: обратное

отображение может оказаться многозначным. Так будет, если какие-либо две точки плоскости (P), как U и V на рис. 224, переходят в одну и ту же точку плоскости $(\overline{P})$, так как такая точка будет иметь по меньшей мере два прообраза.

Если не только само отображение, но и обратное отображение являются однозначными, то рассматриваемое отображение называется *взаимно однозначным* Если же отображение не взаимно однозначное, но не вырождается, то плоскость (P) можно разбить на части, на каждой из которых отображение будет взаимно однозначным

Отображения описываются аналитически с помощью систем координат. Для этого предположим, что в плоскости (P) введена система декартовых координат x, y, а в плоскости $(\overline{P})$ — система координат $\overline{x}$, $\overline{y}$; эти системы также могут быть совмещенными. Тогда если задать координаты x, y точки M, то координаты $\overline{x}$, $\overline{y}$ точки $\overline{M}$ однозначно определяются. Другими словами, отображение определяется зависимостями

$$\overline{x}=\overline{x}(x,\ y);\qquad \overline{y}=\overline{y}(x,\ y). \tag{46}$$

При обратном отображении мы задаем в этих формулах $\overline{x}$ и $\overline{y}$ и находим x и y; для взаимной однозначности отображения нужно, чтобы при этом получалось не более одного решения.

Аналогично определяется отображение трехмерного пространства в трехмерное; такое отображение определяется, взамен (46), формулами

$$\overline{x}=\overline{x}(x,\ y,\ z),\qquad \overline{y}=\overline{y}(x,\ y,\ z),\qquad \overline{z}=\overline{z}(x,\ y,\ z) \tag{47}$$

Можно рассматривать и отображения друг в друга пространств различных размерностей. Например, формулы

$$\overline{x}=\overline{x}(x,\ y),\qquad \overline{y}=\overline{y}(x,\ y)\qquad \overline{z}=\overline{z}(x,\ y)$$

определяют отображение плоскости в трехмерное пространство.

Формулы (X,5) можно рассматривать как формулы, определяющие отображение m-мерного пространства с координатами $t_1,\ t_2,\ \ldots,\ t_m$ в n-мерное пространство с координатами $x_1,\ x_2,\ \ldots,x_n$. При этом система координат не обязана быть декартовой.

14. Нелинейное отображение в малом. Рассмотрим отображение, определенное формулами (46), вблизи некоторой точки $M_0(x_0;\ y_0)$, которая отображается в точку $\overline{M}_0(\overline{x}_0;\ \overline{y}_0)$. Поскольку с точностью до малых высшего порядка приращение любой функции можно заменить ее дифференциалом (п. IX.11), то, пренебрегая этими малыми, получим

$$\left.\begin{aligned} \Delta\overline{x}&=\left(\frac{\partial\overline{x}}{\partial x}\right)_0\Delta x+\left(\frac{\partial\overline{x}}{\partial y}\right)_0\Delta y,\\ \Delta\overline{y}&=\left(\frac{\partial\overline{y}}{\partial x}\right)_0\Delta x+\left(\frac{\partial\overline{y}}{\partial y}\right)_0\Delta y, \end{aligned}\right\} \tag{48}$$

где $\Delta\overline{x}=\overline{x}-\overline{x}_0$, $\Delta x=x-x_0$ и т. д. (можно сказать, что это — декартовы координаты, отсчитываемые соответственно от M_0 и $\overline{M}_0$), а индекс «нуль» указывает на то, что производные берутся в точке M_0 Сравнивая формулы (48) с формулами (27), мы заключаем, что *нелинейное отображение в бесконечно малой окрестности любой точки является* с точностью до малых высшего порядка *линейным*.

На основании п. 6 мы заключаем, что если определитель

$$\begin{vmatrix} \dfrac{\partial \overline{x}}{\partial x} & \dfrac{\partial \overline{x}}{\partial y} \\ \dfrac{\partial \overline{y}}{\partial x} & \dfrac{\partial \overline{y}}{\partial y} \end{vmatrix} = \frac{D(\overline{x}, \overline{y})}{D(x, y)} \qquad (49)$$

(обозначение см. в п. IX.13) в какой-либо точке M_0 не равен нулю, то рассматриваемое отображение в малой окрестности этой точки является взаимно однозначным и даже с точностью до малых высшего порядка аффинным. При этом абсолютное значение определителя (49) равно коэффициенту изменения площадей бесконечно малых фигур при рассматриваемом отображении; этот коэффициент уже не постоянен во всей плоскости, как для линейного отображения, а принимает, вообще говоря, в различных точках разные значения. Это, в частности, дает возможность приписать самостоятельный смысл знаменателю и числителю в правой части (49): они равны площади бесконечно малой фигуры до и после отображения.

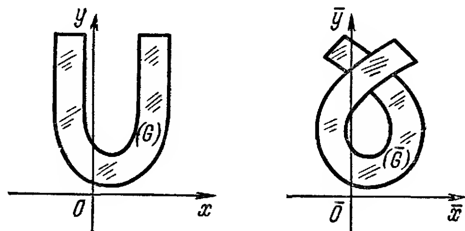


Рис. 225.

Если в какой-либо точке якобиан (49) равен нулю, то *в этой точке* рассматриваемое отображение вырождается, а именно, площадь бесконечно малой фигуры после отображения становится малой высшего порядка. Если же якобиан (49) тождественно равен нулю, то отображение вырождается во всей плоскости, так что при отображении размерность понижается: плоскость отображается на некоторую линию, не обязательно прямую, или даже в точку.

Не следует думать, что если якобиан (49) не равен нулю ни в какой точке некоторой конечной области, то отображение будет в этой области взаимно однозначным Отображение может быть не вырожденным ни в какой точке, но тем не менее не быть взаимно однозначным (см., например, рис. 225).

Аналогичными свойствами обладает отображение трехмерного пространства в трехмерное; это отображение может задаваться формулами (47). Здесь также существенное значение имеет якобиан $\dfrac{D(\overline{x}, \overline{y}, \overline{z})}{D(x, y, z)}$, абсолютное значение которого равно коэффициенту изменения объемов бесконечно малых тел. (Каков геометрический смысл знака якобиана?)

В случае, если этот якобиан тождественно равен нулю, может возникнуть вопрос о «степени» вырождения, т. е. о том, отобразится ли пространство x, y, z на поверхность или на линию (или даже на точку) пространства $\overline{x}, \overline{y}, \overline{z}$. Ответ на этот вопрос вытекает из рассмотрений п. 6 и того, что каждое отображение в бесконечно малом является линейным. Надо рассмотреть матрицу

$$\begin{pmatrix} \dfrac{\partial \overline{x}}{\partial x} & \dfrac{\partial \overline{x}}{\partial y} & \dfrac{\partial \overline{x}}{\partial z} \\ \dfrac{\partial \overline{y}}{\partial x} & \dfrac{\partial \overline{y}}{\partial y} & \dfrac{\partial \overline{y}}{\partial z} \\ \dfrac{\partial \overline{z}}{\partial x} & \dfrac{\partial \overline{z}}{\partial y} & \dfrac{\partial \overline{z}}{\partial z} \end{pmatrix}. \qquad (50)$$

Ее ранг в любой точке меньше трех (почему?). Если он равен двум (в некоторых точках допускается понижение ранга), то пространство x, y, z отображается на двумерную поверхность. Если он нигде не превышает единицы, но не равен нулю тождественно, то отображение происходит на одномерную линию Наконец, если он тождественно равен нулю (это значит, что все элементы матрицы (50) тождественно равны нулю, что будет только для тождественно постоянных $\bar{x}$, $\bar{y}$, $\bar{z}$), то все пространство отображается на точку.

Аналогичная картина получается при отображении пространств любых, не обязательно одинаковых размерностей друг в друга Мы уже говорили, что формулы (X.5) можно рассматривать как формулы, определяющие отображение m-мерного пространства с координатами t_1, t_2, ..., t_m в n-мерное пространство с координатами x_1, x_2, ... x_n Чтобы выяснить, какой размерности многообразие получится в результате этого отображения, нужно составить матрицу размера $n \times m$

$$\begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & \dots & \frac{\partial x_1}{\partial t_m} \\ \frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & \dots & \frac{\partial x_2}{\partial t_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial x_n}{\partial t_1} & \frac{\partial x_n}{\partial t_2} & \dots & \frac{\partial x_n}{\partial t_m} \end{pmatrix}.$$

Если ранг этой матрицы равен k (причем в некоторых точках допускается понижение ранга), то и указанная размерность равна k.

15. Функциональная зависимость функций. Результаты п. 14 можно применить к понятию «функциональной зависимости» системы функций Допустим сначала, что заданы три функции от трех независимых переменных

$$F_1(x, y, z), \quad F_2(x, y, z), \quad F_3(x, y, z). \tag{51}$$

Эти функции называются *функционально зависимыми* друг с другом, если между ними имеет место соотношение вида

$$\Phi(F_1(x, y, z), \quad F_2(x, y, z), \quad F_3(x, y, z)) \equiv 0, \tag{52}$$

где Φ — некоторая функция трех переменных, $\Phi = \Phi(\lambda, \mu, \nu)$, сама по себе не тождественно равная нулю В противном случае функции (51) называются *функционально независимыми* Разрешая соотношение (52) относительно одной из переменных F_1, F_2, F_3, можно сказать также, что функции (51) называются функционально зависимыми, если одна из них является функцией остальных

Например, функции

$$\left(\frac{x-y}{x+z}\right)^2, \quad \ln(x+z), \quad x-y \tag{53}$$

функционально зависимы, так как если их обозначить через F_1, F_2 и F_3, то $F_1 e^{2F_2} - F_3^2 \equiv 0$.

Чтобы установить признак функциональной зависимости в общем случае, рассмотрим вспомогательное отображение

$$\left.\begin{aligned} \lambda &= F_1(x, y, z), \\ \mu &= F_2(x; y, z), \\ \nu &= F_3(x, y, z). \end{aligned}\right\} \tag{54}$$

Для функциональной зависимости системы (51) нужно, чтобы существовало соотношение (52), т. е. соотношение $\Phi(\lambda, \mu, \nu)=0$, справедливое для всех x, y, z. Оно определяет в пространстве λ, μ, ν некоторую поверхность, и мы получаем, что для функциональной зависимости функций (51) нужно, чтобы при отображении (54) пространство x, y, z переходило в поверхность в пространстве λ, μ, ν, т. е. *чтобы это отображение было вырожденным.*

Применяя результат п. 14, приходим к условию

$$\frac{D(F_1, F_2, F_3)}{D(x, y, z)} \equiv 0, \tag{55}$$

необходимому и достаточному для функциональной зависимости функций (51). (Проверьте выполнение условия (55) для функций (53).)

Если же ранг матрицы

$$\begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \\ \frac{\partial F_3}{\partial x} & \frac{\partial F_3}{\partial y} & \frac{\partial F_3}{\partial z} \end{pmatrix}$$

равен единице, то, как мы видели в п. 14, пространство x, y, z при отображении переходит в линию в пространстве λ, μ, ν. Пользуясь уравнениями линии (X.2), заключаем, что в данном случае функции F_1, F_2, F_3 связаны друг с другом двумя независимыми соотношениями вида (52).

Аналогичный результат получается и в том случае, когда число независимых переменных и рассматриваемых функций от них различно. Например, две функции от трех переменных $F_1(x, y, z)$, $F_2(x, y, z)$ будут функционально зависимыми, если при отображении

$$\lambda = F_1(x, y, z), \quad \mu = F_2(x, y, z)$$

пространство x, y, z переходит в линию на плоскости λ, μ (с некоторым уравнением $\Phi(\lambda, \mu)=0$). Согласно п. 14 условие для этого состоит в том, чтобы ранг матрицы

$$\begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{pmatrix}$$

был меньше двух, т. е. чтобы

$$\begin{vmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{vmatrix} \equiv 0, \quad \begin{vmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial z} \end{vmatrix} \equiv 0, \quad \begin{vmatrix} \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{vmatrix} \equiv 0.$$

Условия функциональной зависимости системы из любого числа функций от любого числа аргументов выглядят аналогично Отметим, что если функций больше, чем аргументов, то такая система функций всегда функционально зависима.

ПРИМЕНЕНИЕ ЧАСТНЫХ ПРОИЗВОДНЫХ

§ 1. Скалярное поле

1. Производная по направлению и градиент. Пусть в пространстве задана декартова система координат x, y, z; тогда (см. п. IX.9) стационарное скалярное поле можно рассматривать просто как функцию $u=u(x, y, z)$. (Для нестационарного поля дальнейшие рассмотрения надо проводить в любой, но фиксированный момент времени.) Пусть, кроме того, в пространстве дана некоторая точка M. Пусть линия (L) выходит из точки M по направлению l (рис. 226). Тогда *производной* от u *по* этому *направлению* называется скорость изменения поля в данном направлении, отнесенная к единице длины:

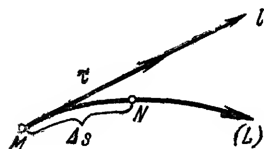


Рис. 226.

$$\frac{\partial u}{\partial l}=\lim_{\Delta s\to 0}\frac{u(N)-u(M)}{\Delta s}. \qquad (1)$$

Для вычисления этой производной допустим, что линия (L) задана в параметрическом виде $\mathbf{r}=\mathbf{r}(s)$, причем за параметр взята длина дуги вдоль (L) (см. п. VII.23). Тогда значение u вдоль (L) будет представлять собой сложную функцию длины дуги: $u(s)=u(x(s), y(s), z(s))$. Искомая производная (1) как раз равна производной $\frac{du}{ds}$, т. е. по правилу (IX. 11) дифференцирования сложной функции

$$\frac{\partial u}{\partial l}=\frac{\partial u}{\partial x}\frac{dx}{ds}+\frac{\partial u}{\partial y}\frac{dy}{ds}+\frac{\partial u}{\partial z}\frac{dz}{ds}.$$

Правую часть удобно представить в виде скалярного произведения двух векторов (см. формулу (VII.12)):

$$\frac{\partial u}{\partial l}=\left(\frac{\partial u}{\partial x}\mathbf{i}+\frac{\partial u}{\partial y}\mathbf{j}+\frac{\partial u}{\partial z}\mathbf{k}\right)\cdot\left(\frac{dx}{ds}\mathbf{i}+\frac{dy}{ds}\mathbf{j}+\frac{dz}{ds}\mathbf{k}\right).$$

Первый из них называется *градиентом* поля u и обозначается

$$\operatorname{grad} u=\frac{\partial u}{\partial x}\mathbf{i}+\frac{\partial u}{\partial y}\mathbf{j}+\frac{\partial u}{\partial z}\mathbf{k}; \qquad (2)$$

его смысл будет показан несколько позже. Второй вектор

$$\frac{dx}{ds}\mathbf{i}+\frac{dy}{ds}\mathbf{j}+\frac{dz}{ds}\mathbf{k}=\frac{d(x\mathbf{i}+y\mathbf{j}+z\mathbf{k})}{ds}=\frac{d\mathbf{r}}{ds}=\boldsymbol{\tau}$$

— это единичный вектор направления l (см. п. VII.23). Таким образом,

$$\frac{\partial u}{\partial l} = \operatorname{grad} u \cdot \boldsymbol{\tau}. \tag{3}$$

Первый множитель в правой части при заданном поле зависит лишь от выбора точки M; второй множитель зависит лишь от направления l. В частности, мы видим, что $\frac{\partial u}{\partial l}$ не зависит от выбора конкретной линии (L) среди всевозможных линий, имеющих в M заданное направление l. (Отметим, кстати, что производная $\partial^2 u/\partial l^2$ уже зависит от такого выбора.)

Согласно формуле (VII.5) из (3) получаем

$$\frac{\partial u}{\partial l} = \text{пр}_l (\operatorname{grad} u) = \operatorname{grad}_l u \tag{4}$$

(это обозначение для проекции градиента на направление l).

Отметим, что производные u'_x, u'_y и u'_z также являются производными по направлению: например, u'_x — это производная по направлению, параллельному оси x.

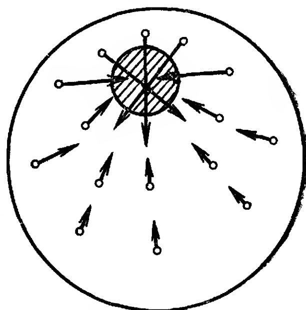


Рис. 227.

Приведем еще полезную формулу, содержащую градиент и основанную на определении (IX.7) полного дифференциала:

$$du = \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy + \frac{\partial u}{\partial z} dz =$$
$$= \left(\frac{\partial u}{\partial x} \mathbf{i} + \frac{\partial u}{\partial y} \mathbf{j} + \frac{\partial u}{\partial z} \mathbf{k}\right) \cdot (dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}) =$$
$$= \operatorname{grad} u \cdot d(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = \operatorname{grad} u \cdot d\mathbf{r}.$$

Пусть заданы поле u и точка M; поставим вопрос: по какому направлению l производная $\partial u/\partial l$ самая большая? Согласно формуле (4) этот вопрос сводится к следующему: на какое направление проекция вектора $\operatorname{grad} u$ самая большая? Очевидно, что любой вектор при проектировании на различные направления дает самую большую проекцию, равную его модулю, при проектировании на его собственное направление. Таким образом, *вектор* $\operatorname{grad} u$ *в точке* M *указывает в сторону самого быстрого возрастания поля «u», причем эта наибольшая скорость в расчете на единицу длины равна* $|\operatorname{grad} u|$; чем быстрее меняется поле, тем этот модуль больше. На рис. 227 показаны векторы градиента температуры в отдельных точках теплопроводящей среды, подогреваемой изнутри, из заштрихованной зоны, и охлаждаемой снаружи. Градиент температуры направлен «к печке».

Полученный физический смысл градиента показывает также, что *градиент инвариантно связан с рассматриваемым полем*, т. е. остается неизменным (инвариантным) при замене декартовых осей. (Этого не было видно из определения градиента (2), данного в «неинвариантной» форме, «привязанной» к какой-то одной системе координат.) Более того, если задано поле u, то в каждой точке пространства можно найти направление и скорость наибыстрейшего возрастания поля u; так можно найти вектор $\operatorname{grad} u$, не прибегая к координатам и к заданию $u(x, y, z)$. Итак, градиент скалярного поля образует вполне определенное векторное поле.

Аналогичное требование инвариантности ставится для всех основных понятий теории векторного поля, которая в этой главе не будет изучаться. Дело в том, что при замене осей декартовой системы координат, хотя векторы остаются неизменными, инвариантными, но их проекции меняются. Таким образом, если какое-либо понятие теории векторного поля сформулировано с помощью координат или проекций этого поля, то надо проверить, что это понятие удовлетворяет требованию инвариантности относительно изменения этих координат и проекций при повороте осей координат.

Покажем применение понятия градиента к вычислению *скорости изменения поля вдоль траектории*. Пусть задано поле u, вообще говоря, нестационарное, т. е. $u = u(x, y, z, t)$. Пусть, далее, задан закон движения $\mathbf{r} = \mathbf{r}(t)$ некоторой частицы M. Если рассматривать значение u в M по мере движения, то это значение будет представлять собой сложную функцию времени: $u = u(x(t), y(t), z(t), t)$ Для вычисления искомой скорости изменения этого значения можно применить преобразования, аналогичные приведенным выше, что даст значение *полной производной*

$$\frac{du}{dt} = \operatorname{grad} u \cdot \mathbf{v} + \frac{\partial u}{\partial t} = \frac{\partial u}{\partial \tau} v + \frac{\partial u}{\partial t},$$

где $\mathbf{v} = \frac{d\mathbf{r}}{dt}$ — вектор скорости движения частицы, а $\frac{\partial u}{\partial \tau}$ — производная по направлению касательной к траектории.

Если поле стационарное $\left(\frac{\partial u}{\partial t} = 0\right)$, то справа остается только первое слагаемое. Таким образом, оно дает скорость изменения поля, полученную только за счет перехода точки M вдоль траектории от одних значений u к другим; например, если u — температура, то за счет перехода из менее нагретой части пространства в более нагретую и т. п. Эта скорость называется *переносной* (*конвекти́вной*). Второе слагаемое дает скорость изменения поля в неподвижной точке, полученную из-за нестационарности поля; эта скорость называется *местной* (*лока́льной*). В общем случае действуют оба указанных фактора и скорость изменения поля вдоль траектории складывается из переносной и местной скоростей изменения поля.

2. Поверхности уровня. Поверхности уровня для поля $u(x, y, z)$ (см. п. IX.7) — это поверхности, на которых поле имеет постоянное значение, $u(x, y, z) = \text{const}$; в зависимости от физического смысла поля они могут называться изотермическими, изобарическими и т. п. поверхностями. Между этими поверхностями и градиентом поля имеется простая связь: *в каждой точке M градиент поля нормален* (т. е. перпендикулярен касательной плоскости) *к поверхности уровня, проходящей через M.*

Действительно (рис. 228), если ΔC мало, то вблизи M поверхности $u=C$ и $u=C+\Delta C$ можно считать почти плоскими и $\frac{\partial u}{\partial l} \approx \frac{\Delta u}{\Delta s} = \frac{\Delta C}{\Delta s}$. Но ясно, что Δs будет самым малым, а потому производная $\frac{\partial u}{\partial l}$ — самой большой, если l направлено по нормали к поверхности. Отсюда и следует наше утверждение.

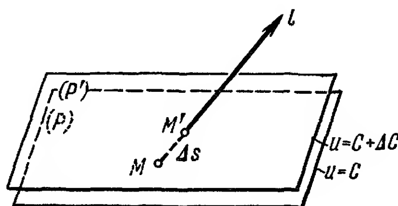


Рис. 228.

Из него, в частности, получаем решение следующей задачи: найти *уравнение касательной плоскости, проведенной в точке* $M_0(x_0;\ y_0;\ z_0)$ *к поверхности* (L), *имеющей уравнение* $F(x, y, z)=0$. Для решения этой задачи рассмотрим на минуту поле в пространстве, определенное по формуле $u=F(x, y, z)$. Тогда (L) будет одной из поверхностей уровня этого поля, так как на ней $u=0$. Но тогда вектор

$$(\operatorname{grad} u)_{M_0} = \left(\frac{\partial F}{\partial x}\right)_0 \mathbf{i} + \left(\frac{\partial F}{\partial y}\right)_0 \mathbf{j} + \left(\frac{\partial F}{\partial z}\right)_0 \mathbf{k}$$

(индекс «нуль» указывает, что рассматриваемые производные берутся в точке M_0) перпендикулярен к искомой касательной плоскости. Согласно п. X.7 (задача 2) получаем уравнение этой плоскости:

$$\left(\frac{\partial F}{\partial x}\right)_0 (x-x_0) + \left(\frac{\partial F}{\partial y}\right)_0 (y-y_0) + \left(\frac{\partial F}{\partial z}\right)_0 (z-z_0) = 0. \qquad (5)$$

Последнее уравнение можно записать в форме $dF=0$; подумайте, как его можно было бы вывести непосредственно.

Уравнение поверхности, к которой надо провести касательную плоскость, может быть задано в форме $z=f(x, y)$. Тогда достаточно это уравнение переписать в виде $z-f(x, y)=0$, обозначить левую часть через $F(x, y, z)$ и применить результат (5): $-\left(\frac{\partial f}{\partial x}\right)_0 (x-x_0) - \left(\frac{\partial f}{\partial y}\right)_0 (y-y_0) + (z-z_0) = 0$, т. е.

$$z-z_0 = \left(\frac{\partial f}{\partial x}\right)_0 (x-x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y-y_0). \qquad (6)$$

Правая часть равна полному дифференциалу df, и мы получаем, таким образом, геометрический смысл полного дифференциала функции двух переменных: он равен приращению третьей координаты касательной плоскости (рис. 229).

Вычислим в качестве примера градиент центрально-симметричного поля $u=f(r)$, где $r=|\mathbf{r}|=\sqrt{x^2+y^2+z^2}$. В этом случае поверхностями уровня

служат сферы с центром в начале координат (почему?). Если взять две сферы, радиусы которых отличаются на dr, то значения функции f на них будут отличаться на df. Поэтому скорость изменения поля поперек поверхностей уровня (т. е. вдоль радиусов) равна $\frac{df}{dr}$, а потому

$$\operatorname{grad} u(r) = \frac{df}{dr}\,\mathbf{r}^0 = \frac{1}{r}\frac{df}{dr}\,\mathbf{r}, \tag{7}$$

где $\mathbf{r}^0 = \mathbf{r}/r$ — орт вектора $\mathbf{r}$. (Получите (7) на основе определения (2).)

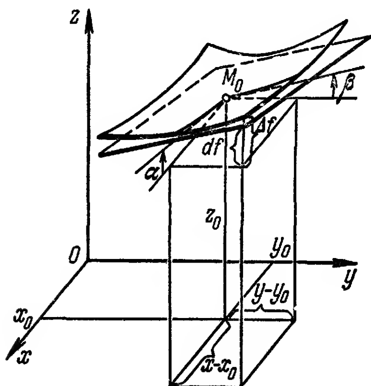


Рис. 229.

$$\operatorname{tg}\varphi = \left(\frac{\partial f}{\partial x}\right)_0, \quad \operatorname{tg}\beta = \left(\frac{\partial f}{\partial y}\right).$$

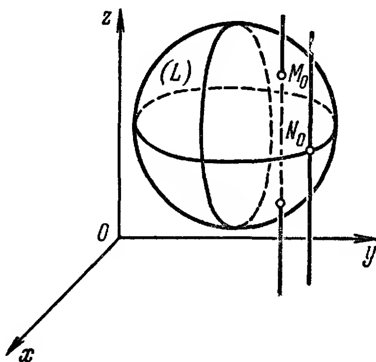


Рис. 230.

3. Неявные функции двух переменных. Неявные функции двух переменных, о которых говорилось в начале п. IX.13, получают теперь новое освещение. Пусть уравнение

$$F(x, y, z) = 0 \tag{8}$$

рассматривается вблизи точки $M_0(x_0; y_0; z_0)$, в которой оно удовлетворяется. Это уравнение определяет в пространстве поверхность (L), проходящую через точку M_0. Если при этом $\left(\frac{\partial F}{\partial z}\right)_0 \neq 0$ [см. условие (IX.16)], то $(\operatorname{grad} F)_{M_0}$ в силу формулы (2) имеет составляющую по оси z. Тогда перпендикулярная к градиенту касательная плоскость к (L) проходит не вертикально, а значит (рис. 230), и прилегающая к ней вблизи M_0 поверхность (L) образует с осью z положительный угол. Поэтому *вблизи M_0 уравнение* (8) *определяет зависимость $z(x, y)$*. Эта зависимость определяется только *локально* (т. е. вблизи M_0 или, как говорят, «*в малом*»), так как если отойти подальше от M_0, то может оказаться, что при некоторых значениях x, y получится несколько значений z или ни одного значения z (см. рис 230). Отметим, что условие существования системы неявных функций, рассмотренное в п. IX.13, также гарантирует существование этих функций, вообще говоря, лишь локально.

Если $\left(\frac{\partial F}{\partial z}\right)_0 = 0$, то касательная плоскость к (L) расположена в рассматриваемой точке вертикально, как в точке N_0. Тогда может получиться, что даже в чрезвычайной близости от N_0 уравнение (8) не определяет однозначной функции $z(x, y)$; так, вблизи N_0 при одних x, y получается два значения z, а при других — ни одного, так как поверхность (L) «заворачивается».

Однако, если в N_0, например, $\frac{\partial F}{\partial y} \neq 0$, то вблизи этой точки уравнение (8) определяет функцию $y(x, z)$. Только в тех точках поверхности (L), в которых

$$\frac{\partial F}{\partial x}=0; \quad \frac{\partial F}{\partial y}=0; \quad \frac{\partial F}{\partial z}=0, \tag{9}$$

может случиться, что уравнение (8) не определяет, даже локально, одну из координат как функцию остальных.

Точки, в которых удовлетворяются условия (9), называются *особыми точками* поверхности (L). Большинство поверхностей не имеет особых точек так как получаемая для их отыскания система четырех уравнений, (8) и (9), с тремя неизвестными, как правило, несовместна. Из наиболее известных поверхностей особой точкой обладают только конические поверхности: у них особая точка в вершине

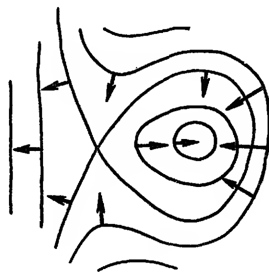


Рис. 231.

4. Плоские поля. Все понятия, введенные для пространственного поля, переносятся с соответствующими упрощениями на плоские поля (см. конец п. IX.9). Так, градиент поля $u(x, y)$, $\operatorname{grad} u = \frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j}$ представляет собой вектор, лежащий в плоскости x, y. В каждой точке градиент поля нормален линии уровня поля, т. е. линии $u(x, y) = \text{const}$, проходящей через эту точку (рис. 231). При этом из смысла градиента вытекает, что его модуль приближенно обратно пропорционален расстоянию между линиями уровня: там, где линии сближаются, градиент длиннее.

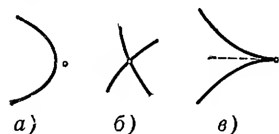


Рис. 232.

а) Изолированная точка; б) точка самопересечения, в) точка возврата

Уравнение касательной к линии

$$f(x, y) = 0 \tag{10}$$

имеет вид (5), но, естественно, без третьего слагаемого.

Уравнение (10) локально определяет функцию $y(x)$, если $f'_y \neq 0$. *Особые точки* на линии (10) — это точки, в которых

$$f'_x = 0 \quad \text{и} \quad f'_y = 0. \tag{11}$$

Если на минуту представить себе поверхность (L) с уравнением $z = f(x, y)$, то линия (10) получается в результате пересечения (L) с плоскостью $z = 0$. Если в какой-либо точке выполнены условия (10) и (11), то из формулы (6) следует, что в этой точке указанная плоскость касается поверхности (L). В п. 9 будет показано, какой вид имеет линия пересечения поверхности со своей касательной плоскостью вблизи точки касания. Оказывается, что обычно особые точки линии являются *изолированными точками*, *точками самопересечения* или, реже, *точками возврата* (рис. 232).

5. Огибающая однопараметрического семейства линий. Пусть рассматривается семейство линий, зависящих от одного параметра C (*однопараметрическое семейство*). Уравнение этих линий можно записать в общем виде.

$$F(x, y; C) = 0. \tag{12}$$

причем при каждом конкретном C получается индивидуальная линия из семейства. Часто бывает, что в некоторых частях плоскости эти линии расположены наподобие того, как показано на рис. 233; в этом случае семейство имеет *огибающую*, т. е. *линию* (как правило, в семейство не входящую), *каждой точки которой касается одна из линий семейства*. Чтобы найти уравнение огибающей, заметим, что каждая ее точка принадлежит одной из линий семейства, т. е. в каждой такой точке удовлетворяется уравнение (12), однако величина C будет по мере движения по огибающей меняться, $C=C(x)$. Дифференцируя по x равенство (12) «вдоль огибающей», получим

Рис. 233.

$$F_x+F'_y\, y'_{\text{огиб}}+F'_C C'_x=0. \tag{13}$$

Здесь $y'_{\text{огиб}}$ — угловой коэффициент огибающей в ее произвольной точке M. Но в этой точке огибающей касается одна из линий семейства, которая, следовательно, имеет такой же угловой коэффициент, т. е. $y'_{\text{линии}}=y'_{\text{огиб}}$; $y'_{\text{линии}}$ находится из уравнения (12) дифференцированием при фиксированном C:

$$F_x+F'_y\, y'_{\text{линии}}=0. \tag{14}$$

Из (13) и (14) видим, что $F'_C C'_x=0$, но, как было сказано, величина $C(x)$ переменная, т. е. $C'_x\neq 0$, а потому

$$F'_C(x,\ y;\ C)=0. \tag{15}$$

Итак, для точек огибающей наряду с (12) выполняется соотношение (15). Исключая из этих двух уравнений C, получим уравнение огибающей.

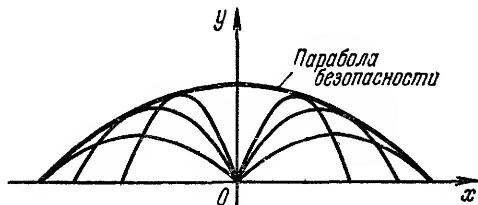


Рис. 234.

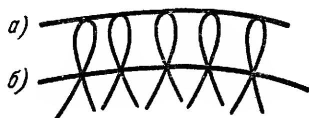


Рис. 235.

а) Огибающая; б) линия точек самопересечения.

Пример. Рассмотрим семейство траекторий полета снаряда в предположениях примера 1 п. II.6 при заданной начальной скорости v_0 и различных углах бросания α. Здесь параметром семейства служит α, и поэтому для нахождения огибающей (рис. 234) продифференцируем уравнение семейства [уравнение (II.11)] по α:

$$0=\frac{x}{\cos^2\alpha}-\frac{gx^2\sin\alpha}{v_0^2\cos^3\alpha}.$$

Если отсюда выразить $\operatorname{tg}\alpha$ и подставить его в уравнение семейства, то получится уравнение огибающей

$$y=\frac{v_0^2}{2g}-\frac{g}{2v_0^2}x^2.$$

Значит, огибающей служит парабола, называемая *параболой безопасности* (почему?).

Приведем еще пример. Так как все нормали к эвольвенте касаются эволюты (п. VII.26), то *эволюта является огибающей семейства всех нормалей к эвольвенте*. Отсюда вытекает способ приближенного построения эволюты: надо построить несколько нормалей к эвольвенте, а затем навести их огибающую.

Следует иметь в виду, что если линии рассматриваемого семейства имеют особые точки (см. п. 4), то, исключая C из (12) и (15), мы наряду с огибающей получим *линию особых точек* (рис. 235). Действительно, как мы видели в п. 4, для таких точек $F'_x = F'_y = 0$, а потому из (13) вытекает (15), даже если $y'_{\text{линии ос. точек}} \neq y'_{\text{линии семейства}}$.

§ 2. Экстремум функции нескольких переменных

6. Формула Тейлора для функции нескольких переменных. Рассмотрим для определенности функцию $f(x, y)$ двух переменных, хотя аналогичные результаты справедливы для любого числа переменных. Оказывается, что формула (IV.62) без какого-либо изменения остается справедливой и для такой функции f.

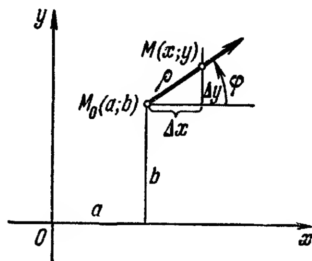


Рис. 236.

Для доказательства выберем произвольное направление, выходящее из точки $(a; b)$ в плоскости аргументов и указанное на рис. 236 стрелкой. Вдоль этого направления функция f зависит от одного аргумента ρ, т. е. $f(x, y) = f^*(\rho)$. К функции f^* можно применить формулу (IV. 62). При этом $\Delta f^* = \Delta f$, но при выяснении связи между дифференциалами функций f^* и f надо учесть следующее. Так как (рис. 236)

$$x = a + \rho\cos\varphi; \quad y = b + \rho\sin\varphi \quad (a,\ b,\ \varphi = \text{const}), \tag{16}$$

то

$$f^*(\rho) = f(x, y) = f(a + \rho\cos\varphi,\ b + \rho\sin\varphi).$$

Таким образом, если при составлении df, d^2f, ... переменные x и y считаются независимыми, то при составлении df^*, d^2f^*, ... они считаются зависящими от ρ. Как известно (см. пп. IX.12 и IX.16), для дифференциала первого порядка это несущественно, т. е. $df^* = df$, но для последующих дифференциалов это, вообще говоря, существенно. Однако в рассматриваемом случае из формул (16) видим, что

$$d^2x = d^3x = \ldots = 0; \quad d^2y = d^3y = \ldots = 0.$$

Поэтому из формул (IX.23) и (IX.24) следует, что $d^2f^* = d^2f$ и аналогично $d^3f^* = d^3f$ и т. д.

Таким образом, из формулы (IV.62) для функции f^* автоматически вытекает справедливость этой же формулы для функции $f(x, y)$.

В применениях эту формулу обычно обрывают, оставляя лишь один или два члена. Тогда получается (см. п. IX.16)

$$f(a+h,\ b+k)=f(a,b)+f'_x(a,\ b)\,h+f'_y(a,\ b)\,k+$$
$$+\text{члены не менее второго порядка малости (по сравнению с } h \text{ и } k); \tag{17}$$

$$f(a+h,\ b+k)=f(a,b)+f_x(a,\ b)\,h+f_y(a,\ b)\,k+$$
$$+\frac{1}{2}[f''_{xx}(a,\ b)\,h^2+2f'_{xy}(a,\ b)\,hk+f''_{yy}(a,\ b)\,k^2]+$$
$$+\text{члены не менее третьего порядка малости.} \tag{18}$$

Как и для функций одного переменного, формулами (17) и (18) можно пользоваться, если $|h|$ и $|k|$ малы, в противном случае формулы могут привести к ошибочным выводам. Во всех случаях применения формулы Тейлора, конечно, предполагается, что рассматриваемые производные существуют и конечны.

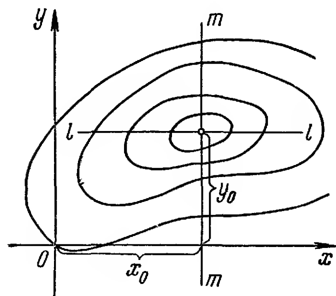


Рис. 237.

7. Экстремум. Как и в п. 6, мы будем для простоты считать, что рассматривается функция двух переменных $z=f(x,\ y)$. Определение *экстремума* дается аналогично случаю функций одного переменного (п. IV.18). Например, мы говорим, что функция имеет *максимум* «в точке» (т. е. при значениях) $x=x_0$, $y=y_0$, если значение $f(x_0,\ y_0)$ больше всех «соседних» значений функции f, т. е. значений $f(x,\ y)$ при x, достаточно близких к x_0, и y, достаточно близких к y_0.

В этом пункте мы будем рассматривать только *внутренние экстремумы*, т. е. точки экстремума, лежащие внутри области определения функции f, причем будем считать, что сама функция f и ее частные производные не имеют разрывов. Примерный вид семейства линий уровня (см. п. IX.1) около точки экстремума показан на рис. 237.

Легко вывести необходимое условие экстремума: если зафиксировать $y=y_0$ и изменять x, т. е. на рис. 237 следовать вдоль прямой ll, то функция f будет иметь при $x=x_0$ экстремум. Но при таком рассмотрении функция $f=f(x,\ y_0)$ будет зависеть только от x и потому согласно п. IV.18 $f'_x(x_0,\ y_0)=0$; производная частная, так как она берется при фиксированном y. Аналогично рассматривается случай следования по прямой $x=x_0$, и мы получаем *необходимые условия экстремума*:

$$f'_x(x_0,\ y_0)=0, \quad f'_y(x_0,\ y_0)=0 \tag{19}$$

(для функций большего числа переменных надо приравнять нулю *все* частные производные первого порядка). Точка (в плоскости аргу-

ментов), в которой выполняются условия (19), называется *стационарной точкой* функции f. Таким образом, в предположениях, указанных в предыдущем абзаце, *все точки экстремума функции f являются ее стационарными точками.*

Обратно, пусть у функции $f(x, y)$ найдена какая-либо стационарная точка $(x_0; y_0)$; будет ли она точкой экстремума? Если в какой-либо области стационарная точка имеется только одна, а существование там экстремума вытекает из физических или каких-либо иных соображений, то ясно, что ответ утвердительный. В других случаях приходится обращаться к достаточным условиям экстремума, к которым мы и переходим.

Как известно из п. IV.18, для функции $f(x)$ одного переменного необходимое условие $f'(x_0)=0$ экстремума является «почти достаточным»: например, если $f''(x_0)\neq 0$, то в точке $x=x_0$ обязательно будет экстремум. Можно было бы ожидать, что и для функции двух переменных при выполнении условия (19) экстремум в точке $(x_0; y_0)$ обязательно будет, если в ней частные производные второго порядка отличны от нуля. Замечательно, что это, вообще говоря, не обязательно: так, функции нескольких переменных доставляют случаи принципиально нового типа.

Так, «график» функции $z=f(x, y)=x^2+y^2$ показан на рис. 220. Условия (19) дают единственную стационарную точку (0; 0). Ясно, что в ней будет минимум, так как $z(0, 0)=0$, а в остальных точках $z>0$. Принципиально иной случай будет для функции $z=-x^2+y^2$, «график» которой показан на рис. 221. И здесь единственной стационарной точкой служит начало координат. При $x=0$ получаем $z=y^2$, т. е. от начала координат вдоль оси y функция в обе стороны возрастает, а в самом начале имеет минимум. Если же $y=0$, то $z=-x^2$, т. е. вдоль оси x функция в обе стороны убывает, а в самом начале имеет максимум. Если рассматривать другие прямые, проходящие через начало координат, то вдоль одних из них функция имеет в начале максимум, а вдоль других — минимум. Такой случай называется *минима́кс*, и здесь экстремума в начале координат не будет, хотя необходимые условия (19) выполняются и частные производные второго порядка не все равны нулю.

Перейдем от примеров к функциям общего вида. Допустим, что в некоторой точке $(x_0; y_0)$ выполняются условия (19), и мы хотим выяснить, действительно ли функция f имеет в этой точке экстремум. Тогда следует воспользоваться формулой Тейлора (18), положив в ней $a=x_0$, $b=y_0$; получим

$$\Delta f=f(x_0+h, y_0+k)-f(x_0, y_0)=$$
$$=\frac{1}{2}[f''_{xx}(x_0, y_0)h^2+2f''_{xy}(x_0, y_0)hk+f''_{yy}(x_0, y_0)k^2]+$$
$$+\text{члены не менее третьего порядка малости,}$$

члены первого порядка отсутствуют из-за условий стационарности

(19). Так как при малых $|h|$, $|k|$ члены третьего порядка значительно меньше членов второго порядка, то знак всей правой части определяется знаком суммы членов второго порядка малости, т. е. знаком квадратичной формы

$$P(h,\ k)=f''_{xx}(x_0,\ y_0)\,h^2+2f''_{xy}(x_0,\ y_0)\,hk+f''_{yy}(x_0,\ y_0)\,k^2; \qquad (20)$$

мы не пишем коэффициент $\frac{1}{2}$, так как он для выяснения знака несуществен. Таким образом, *если эта сумма положительна при всех* h, k (конечно, кроме значений $h=k=0$, когда она равна нулю), *то в точке* $(x_0;\ y_0)$ *будет минимум*, так как тогда при достаточно малых $|h|$, $|k|$ будет $\Delta f>0$, т. е. $f(x_0+h,\ y_0+k)>f(x_0,\ y_0)$. *Если эта сумма отрицательна, то в точке* $(x_0;\ y_0)$ *будет максимум. Если эта сумма может принимать значения обоих знаков, то в точке* $(x_0;\ y_0)$ *будет минимакс и экстремума не будет.* Единственный случай, когда по сумме членов второго порядка нельзя судить о наличии экстремума, тот, когда эта сумма знака менять не может, но может обращаться в нуль (в частности, если она полностью отсутствует). Тогда формулу Тейлора надо продолжить, выписав члены третьего порядка, и провести аналогичное исследование суммы этих членов при тех значениях h, k, при которых сумма членов второго порядка равна нулю. Мы не будем здесь этого делать.

Полученные выводы формулируются совершенно аналогично для функций любого числа переменных. Однако для функций двух переменных легко пойти дальше и получить достаточные признаки экстремума, выраженные непосредственно через значения производных второго порядка в точке $(x_0;\ y_0)$. Для этого вынесем в правой части (20) h^2 за скобки и обозначим $\frac{k}{h}=t$. Тогда получим

$$P(h,\ k)=[(f''_{xx})_0+2\,(f''_{xy})_0\,t+(f''_{yy})_0\,t^2]\,h^2 \qquad (21)$$

(индекс «нуль» указывает на то, что значения производных берутся в рассматриваемой стационарной точке $(x_0;\ y_0)$). Из элементарной алгебры известно, что если дискриминант

$$(f''_{xy})_0^2-(f''_{xx})_0\,(f''_{yy})_0>0, \qquad (22)$$

то многочлен относительно t, стоящий в квадратных скобках, имеет два вещественных нуля, при переходе через которые он меняет знак. Значит, это — случай *минимакса.* Если же

$$(f''_{xy})_0^2-(f''_{xx})_0\,(f''_{yy})_0<0,$$

то указанный многочлен имеет мнимые нули и потому знака не меняет (почему?). Значит, это — случай экстремума. Чтобы узнать, какой именно знак имеет правая часть (21), положим $t=0$. Мы видим, что если

$$(f''_{xy})_0^2-(f''_{xx})_0\,(f''_{yy})_0<0,\quad (f''_{xx})_0>0, \qquad (23)$$

то правая часть (21) положительна при всех t и потому в силу предыдущего абзаца функция f имеет в точке $(x_0; y_0)$ *минимум*. Если же

$$(f''_{xy})_0^2-(f''_{xx})_0(f''_{yy})_0<0, \quad (f''_{xx})_0<0, \tag{24}$$

то функция f имеет *максимум*. Наконец, если

$$(f''_{xy})_0^2-(f''_{xx})_0(f''_{yy})_0=0, \tag{25}$$

то многочлен (21) имеет двойной корень, а потому знака не меняет, но может обратиться в нуль; это — *неопределенный случай*.

Условие положительности квадратичной формы (20) можно вывести также из общей теории квадратичных форм (п. XI.11). Согласно этой теории после некоторого поворота осей координат форма (20) приобретает «диагональный вид»

$$P=\lambda_1 h'^2+\lambda_2 k'^2, \tag{26}$$

где λ_1 и λ_2 — корни характеристического уравнения

$$\begin{vmatrix} (f''_{xx})_0-\lambda & (f''_{xy})_0 \\ (f''_{xy})_0 & (f''_{yy})_0-\lambda \end{vmatrix}=0; \tag{27}$$

здесь h' и k' — приращения координат после поворота. Из (27) следует (проверьте!), что

$$\lambda_1\lambda_2=(f''_{xx})_0(f''_{yy})_0-(f''_{xy})_0^2, \quad \lambda_1+\lambda_2=(f''_{xx})_0+(f''_{yy})_0.$$

Из этих равенств легко вывести, что мы предоставляем читателю, что в случаях (22)—(25) будет соответственно $\lambda_1\lambda_2<0$, $\lambda_1>0$, $\lambda_2>0$, $\lambda_1<0$, $\lambda_2<0$; $\lambda_1\lambda_2=0$ Отсюда с помощью равенства (26) получаем те же выводы, что в предыдущем абзаце.

Для функции $f(x_1, x_2, \ldots, x_n)$ от любого числа n аргументов в стационарной точке взамен (20) надо рассмотреть квадратичную форму

$$\sum_{i,j=1}^{n} (f''_{x_i x_j})_0 h_i h_j, \tag{28}$$

а взамен (27) — уравнение

$$\det(\mathbf{A}-\lambda\mathbf{I})=0, \text{ где } \mathbf{A}=((f''_{x_i x_j})_0)_{nn}. \tag{29}$$

Если все корни уравнения (29) положительны, то форма (28) принимает только положительные значения; такая форма называется *положительно определенной*. В этом случае функция f имеет в рассматриваемой стационарной точке минимум Если все корни уравнения (29) отрицательны, то форма (28) *отрицательно определенная* и функция f имеет максимум Если же уравнение (29) имеет корни обоих знаков, то функция f имеет минимакс.

8. Метод наименьших квадратов. В качестве примера на экстремум функции двух переменных рассмотрим *метод наименьших квадратов* при построении эмпирических формул. Он применяется, если точность грубого метода, указанного в п. I.30, нас не устраивает, а также при автоматизации вычислений. Мы остановимся здесь только на выборе линейной зависимости в случае одной независимой переменной. При этом рассуждают так (для других зависимостей

можно рассуждать аналогично): искомая функция имеет вид $y = kx + b$, однако значения параметров k и b пока неизвестны. При подстановке $x = x_i$ по формуле должно было бы получиться $kx_i + b$, а в результате эксперимента получилось y_i, т. е. формула дает расхождение $y_i - kx_i - b$ с экспериментом, полученное за счет ошибок эксперимента и вычислений, неточной линейности изучаемой зависимости и т. п. Эта разность между левой и правой частями формулы называется *невязкой*. И вот k и b подбирают так, *чтобы сумма квадратов этих невязок*, т. е. $S = \sum_{i=1}^{n} (y_i - kx_i - b)^2$, *была минимально возможной*. (Можно брать и сумму других четных степеней или, например, сумму абсолютных величин невязок, но тогда вычисления будут значительно сложнее. Однако руководствоваться суммой самих невязок, конечно, нельзя, так как она может получиться малой при больших по абсолютной величине слагаемых разного знака.) Мы приходим к задаче на минимум функции $S = S(k, b)$. Пользуясь необходимыми условиями (19), видим, что для минимума должно быть

$$S'_k = -\sum_{i=1}^{N} 2(y_i - kx_i - b)\,x_i = 0; \qquad S'_b = -\sum_{i=1}^{N} 2(y_i - kx_i - b) = 0,$$

откуда

$$k\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i; \quad k\sum_{i=1}^{N} x_i + bN = \sum_{i=1}^{N} y_i.$$

Для нахождения k и b получилась простая система двух уравнений первой степени с двумя неизвестными, поскольку все x_i и y_i даны, которую легко решить. То, что здесь получится действительно минимум S, вытекает из смысла решаемой задачи.

Подобный метод применяется к подбору эмпирических формул другого вида, а также и к иным задачам.

Рассмотрим такой пример. Пусть для определения некоторых величин x, y эмпирически выяснялись соотношения между ними, причем оказалось, что приближенно

$$\left.\begin{aligned} x + y &= 5{,}8, \\ x + 2y &= 8{,}1, \\ 2x + 3y &= 13{,}2. \end{aligned}\right\} \qquad (30)$$

С виду эта система противоречивая: сложив первые два уравнения, вступаем в противоречие с третьим. Однако возможны ведь ошибки эксперимента! Поэтому постараемся удовлетворить системе (30) как можно точнее, чтобы сумма квадратов невязок оказалась как можно меньшей. Значит, надо найти такие x, y, для которых величина

$$S = (x + y - 5{,}8)^2 + (x + 2y - 8{,}1)^2 + (2x + 3y - 13{,}2)^2$$

минимальна. Применение необходимого условия (19) дает

$$S'_x=2(x+y-5{,}8)+2(x+2y-8{,}1)+2(2x+3y-13{,}2)\,2=0;$$
$$S'_y=2(x+y-5{,}8)+2(x+2y-8{,}1)\,2+2(2x+3y-13{,}2)\,3=0.$$

Сокращая на 2, получим

$$\left.\begin{aligned} 6x+9y&=5{,}8+8{,}1+2\cdot 13{,}2=40{,}3,\\ 9x+14y&=5{,}8+2\cdot 8{,}1+3\cdot 13{,}2=61{,}6,\end{aligned}\right\}$$

откуда, решая простейшим способом, получим $x=3{,}3$, $y=2{,}3$. Конечно, система (30) удовлетворяется этими значениями лишь приближенно. Естественно, чем больше соотношений вида (30) между x и y будет определено, тем достовернее получатся значения x и y, если нет систематических ошибок эксперимента, так как при этом случайные ошибки в отдельных соотношениях погашают друг друга Аналогично решаются другие системы приближенных, в частности эмпирических, уравнений, если число уравнений больше, чем число неизвестных.

Метод наименьших квадратов, открытый французским математиком А. Лежа́ндром (1752—1833) и Гауссом, в настоящее время имеет многочисленные приложения.

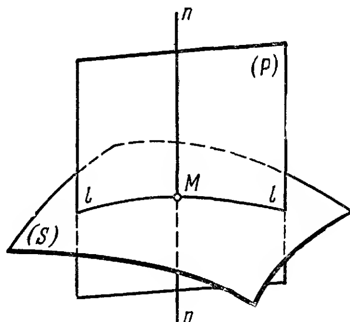


Рис. 238.

9. Кривизна поверхностей. Классификация стационарных точек, указанная в п. 7, непосредственно связана с классификацией поверхностей.

Рассмотрим произвольную поверхность (S) и на ней некоторую точку M (рис. 238). Если провести через эту точку нормаль nn к поверхности, а затем через эту нормаль произвольную плоскость (P), то эта плоскость пересечет (S) по некоторой линии ll, *нормальному сечению*, которое имеет в точке M определенную кривизну k (п. VII.24). Если теперь поворачивать плоскость (P) вокруг нормали nn, то нормальное сечение будет меняться и поэтому величина k тоже, вообще говоря, будет меняться. Чтобы выяснить закон этого изменения, выберем систему декартовых координат так, чтобы начало координат находилось в точке M, а ось z пошла по нормали nn. Тогда вблизи M уравнение поверхности (S) можно представить в виде $z=z(x, y)$, причем значение $z(0, 0)$ будет стационарным (почему?). Рассуждая, как в конце п. 7, мы получим, что после некоторого поворота осей координат вокруг nn уравнение (S) примет вид

$$z=\frac{1}{2}(\lambda_1 x'^2+\lambda_2 y'^2)+\text{члены высшего порядка малости.} \tag{31}$$

Пусть плоскость (P) образует с плоскостью $x'Mz$ угол φ; тогда, переходя к полярным координатам, получим $x'=\rho\cos\varphi$, $y'=\rho\sin\varphi$, откуда $z=\frac{1}{2}(\lambda_1\cos^2\varphi+\lambda_2\sin^2\varphi)\rho^2+\ldots$, т. е. в точке M будет $\frac{dz}{d\rho}=0$, $\frac{d^2z}{d\rho^2}=\lambda_1\cos^2\varphi+\lambda_2\sin^2\varphi$. По формуле (VII.37) получаем искомое выражение для кривизны $k=|\lambda_1\cos^2\varphi+\lambda_2\sin^2\varphi|$. При этом возможны следующие случаи (*три типа точек у поверхностей*):

1. Пусть $\lambda_1\lambda_2>0$, т. е. λ_1 и λ_2 имеют одинаковый знак. Тогда все нормальные сечения вблизи M выгнуты в одну сторону, причем величина k заключена между $|\lambda_1|$ и $|\lambda_2|$, причем $k=|\lambda_1|$ для $\varphi=0$, т. е. для плоско-

сти $x'Mz$, и $k=|\lambda_2|$ для $\varphi=\frac{\pi}{2}$, т. е. для плоскости $y'Mz$ (это — *главные нормальные сечения*). Такая точка M называется *эллиптической* (а при $|\lambda_1|=|\lambda_2|$ — *сферической*) точкой поверхности (S), например, у эллипсоида или у двуполостного гиперболоида все точки эллиптические. Как видно из уравнения (31), касательная плоскость к (S) в точке M имеет вблизи M с (S) только одну общую точку M, параллельные ей плоскости сначала дают в пересечении с (S) бесконечно малый эллипс, оси которого расположены по главным нормальным сечениям, а затем форма сечения может стать более сложной (рис. 239, *а*).

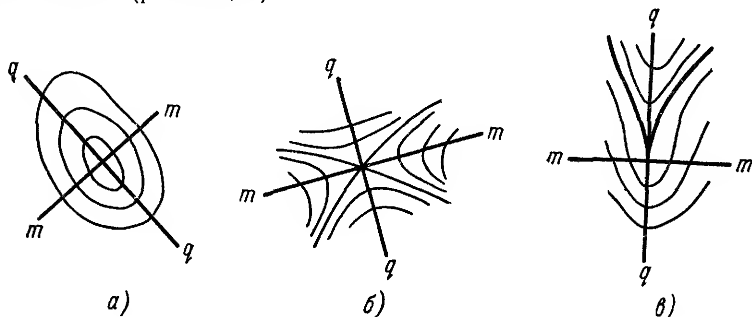


Рис. 239.
mm и qq — главные нормальные сечения.

2. Пусть $\lambda_1\lambda_2<0$. Тогда некоторые нормальные сечения выгнуты вблизи M в одну сторону и имеют в M положительную кривизну, а некоторые — в другую сторону и также имеют в M положительную кривизну; так будет, например, для точек однополостного гиперболоида. Среди первых нормальных сечений одно имеет наибольшую кривизну $|\lambda_1|$, а среди вторых — одно имеет наибольшую кривизну $|\lambda_2|$. Эти главные нормальные сечения также взаимно перпендикулярны. Точка M называется *гиперболической*. Касательная плоскость к (S) в точке M пересекает (S) по двум линиям, пересекающимся в M под положительным углом, бесконечно близкие параллельные к ней плоскости дают в пересечении в бесконечной близости M гиперболы с осями по главным нормальным сечениям (рис. 239, *б*).

3. Пусть $\lambda_1\lambda_2=0$. Тогда если λ_1 и λ_2 не оба равны нулю, то все нормальные сечения выгнуты вблизи M в одну сторону и имеют в M положительную кривизну, но одно имеет нулевую кривизну в M, а перпендикулярное сечение имеет в M наибольшую кривизну. Точка называется *параболической*; например, у цилиндрической или конической поверхности все точки параболические. Типичная картина пересечений (S) с плоскостями, параллельными касательной плоскости в M, показана на рис. 239, *в*, но довольно часто бывают и другие картины. К этому же типу принадлежит и тот случай, когда $\lambda_1=\lambda_2=0$, т. е. $k=0$ при всех φ, такие точки называются *точками уплощения* поверхности (S), так как ясно, что у плоскости все точки будут такими.

В приведенных примерах поверхностей все точки имели одинаковый тип. Но это необязательно, например, у *тора*, т. е. бублика идеальной формы, имеются точки всех трех типов (где они?).

Во всех случаях произведение $\lambda_1\lambda_2$ называется *полной* (*гауссовой*) *кривизной поверхности* (S) *в точке* M. Полная кривизна обладает тем замечательным свойством, что когда поверхность изгибается без растяжений, то эта кривизна не меняется. Например, если лист бумаги изогнуть произволь-

ным способом, то полученная поверхность будет иметь в каждой своей точке нулевую полную кривизну. По этой же причине никакой кусок сферы нельзя разложить на плоскости без деформаций, т. е. невозможны географические карты без искажений.

Поверхность может иметь *особые точки* (ими обычно служат или *изолированные точки*, или *«конические» точки*, как вершина у кругового конуса, хотя бывают особые точки и гораздо более сложного вида), а также целые *особые линии* (это чаще всего *изолированные линии* или *линии самопересечения*; бывают также *«ребра возврата»* и другие виды). Часто рассматриваются *поверхности «с краем»*, т. е. вырезанные из более полной поверхности наподобие куска плоскости или куска сферы.

10. Условный экстремум. В задачах, рассмотренных в п. 7, независимые переменные не были связаны между собой никакими соотношениями; такой экстремум называется *безусловным*. Встречаются также задачи на *условный* экстремум, в которых независимые переменные связаны между собой соотношениями, имеющими вид равенств. Начнем с функций двух переменных.

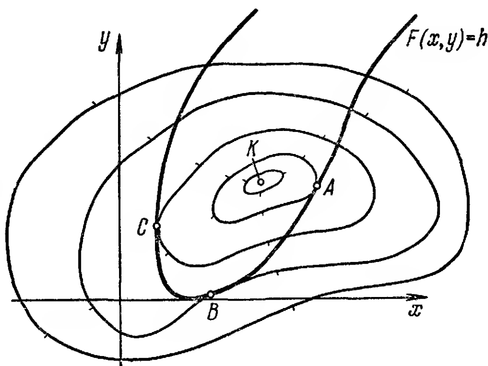


Рис. 240.

Пусть ищется максимум или минимум функции $z=f(x, y)$ *при условии*, что переменные x и y связаны соотношением

$$F(x, y)=h. \qquad (32)$$

Это означает, что значения функции f рассматриваются и сравниваются только для точек (в плоскости аргументов), лежащих на линии с уравнением (32). Например, на рис. 240, где изображены линии уровня некоторой функции f, безусловный максимум достигается в точке K, тогда как условных экстремумов три: два максимума в точках A и C и один минимум в точке B (почему?). Безусловный максимум — это как бы вершина горы, а условный — это самая высокая точка горной тропы, проекция которой на плоскость x, y имеет уравнение (32).

Если из *уравнения связи* (32) можно выразить y через x, то этот результат можно подставить в выражение для z

$$z=f[x, y(x)] \qquad (33)$$

и получить, таким образом, z как функцию одного независимого переменного. Так как условия больше нет, оно учтено подстановкой $y=y(x)$, то задача об отыскании экстремума z становится задачей на безусловный экстремум. Аналогичный результат получится, если

уравнение (32) можно разрешить относительно x или если уравнение линии (32) можно представить в параметрическом виде.

Однако такое разрешение уравнения (32) не всегда возможно и целесо образно Тогда можно рассуждать так. Уравнение связи (32) определяет принципиально некоторую зависимость $y=y(x)$, хотя бы нам явно и не известную. Таким образом, z является сложной функцией (33) независимой переменной x, и необходимое условие экстремума дает по формуле производной сложной функции

$$\frac{dz}{dx}=f'_x+f'_y\frac{dy}{dx}=0. \tag{34}$$

Здесь $\frac{dy}{dx}$ означает производную неявной функции $y(x)$, определенной из условия (32). Значит (см. п. IX.13),

$$F'_x+F'_y\frac{dy}{dx}=0, \qquad \frac{dy}{dx}=-\frac{F'_x}{F'_y}.$$

Подставляя это выражение в (34), видим, что в точке условного экстремума

$$f'_x-\frac{F'_x}{F'_y}f'_y=0, \quad \text{т. е.} \quad \frac{f'_x}{F'_x}=\frac{f'_y}{F'_y}.$$

Обозначим значение последнего отношения в рассматриваемой точке через $-\lambda$ ($-$ взят для удобства, само λ может иметь любой знак). Тогда получим в точке условного экстремума

$$\frac{f'_x}{F'_x}=\frac{f'_y}{F'_y}=-\lambda, \tag{35}$$

т е.

$$f'_x+\lambda F'_x=0; \quad f'_y+\lambda F'_y=0. \tag{36}$$

Обозначим

$$f^*(x,\ y,\ \lambda)=f(x,\ y)+\lambda F(x,\ y), \tag{37}$$

где λ—неизвестный параметр, называемый *множителем Лагранжа*, который разработал данный метод. Тогда уравнения (36) можно записать в виде

$$f^{*\prime}_x=0, \quad f^{*\prime}_y=0. \tag{38}$$

Таким образом, получаются такие же уравнения, как и в случае безусловного экстремума [см. уравнения (19)], однако они составляются не для самой функции f, а для измененной функции f^*, определенной по формуле (37). Уравнения (38) вместе с уравнением связи (32) образуют систему трех уравнений с тремя неизвестными x, y, λ, из этих уравнений и находятся возможные точки условного экстремума. Полученные условия являются только необходимыми, достаточные признаки условного экстремума можно аналогичным образом вывести из достаточных признаков безусловного экстремума (для чего, например, подсчитать d^2z/dx^2), но мы этого делать здесь не будем.

Множитель Лагранжа λ имеет простой смысл. Для выяснения его обозначим координаты точки условного экстремума и само экстремальное значение соответственно через $\bar{x}$, $\bar{y}$ и $\bar{z}$. До сих пор мы считали h зафиксированным, но если менять h, то эти три величины будут зависеть от h. Так как $\bar{z}(h)=f(\bar{x}(h),\ \bar{y}(h))$, то

$$\frac{d\bar{z}}{dh}=f'_x\frac{d\bar{x}}{dh}+f'_y\frac{d\bar{y}}{dh}. \tag{39}$$

С другой стороны, в силу (32)

$$F'_x \frac{d\bar{x}}{dh} + F'_y \frac{d\bar{y}}{dh} = 1. \tag{40}$$

Из (39), (35) и (40) легко получаем, что $\frac{d\bar{z}}{dh} = -\lambda$. Итак, множитель λ, с точностью до знака, равен скорости изменения экстремального значения $\bar{z}$ при изменении параметра h в условии.

Для функций любого числа переменных и для любого числа связей исследование условного экстремума проводится аналогично. (Напомним, что согласно п. X.2 число связей должно быть меньше числа переменных.)

Например, если ищется экстремум функции $f(x, y, z, u, v)$ при условиях

$$F_1(x, y, z, u, v) = 0, \quad F_2(x, y, z, u, v) = 0, \quad F_3(x, y, z, u, v) = 0, \tag{41}$$

то надо поступать так, как если бы надо было искать безусловный экстремум для функции

$$f^* = f + \lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3,$$

где λ_1, λ_2, λ_3 — неизвестные множители Лагранжа. Требование стационарности для f^* дает

$$f^{*\prime}_x = 0; \quad f^{*\prime}_y = 0; \quad f^{*\prime}_z = 0; \quad f^{*\prime}_u = 0; \quad f^{*\prime}_v = 0,$$

что вместе с (41) дает 5+3 уравнений с 5+3 неизвестными x, y, z, u, v, λ_1, λ_2, λ_3.

Решение задачи на условный экстремум лежит в основе одного из распространенных методов численного нахождения безусловных экстремумов, именно, *метода наискорейшего спуска.* Мы опишем один из вариантов этого метода на примере функции двух переменных, хотя в принципе он применим для любого числа переменных. Пусть требуется найти точку минимума функции $f(x, y)$. Оказывается, что для сколько-нибудь сложной функции f решить систему уравнений (19) довольно трудно, да к тому же при этом затрачивается лишняя работа на отыскание стационарных точек, не являющихся точками минимума. Поэтому можно применить следующий итерационный метод. Начинаем с какого-либо нулевого приближения $M_0(x_0, y_0)$. В этой точке функция убывает быстрее всего по направлению $-\operatorname{grad} f = -f'_x(x_0, y_0)\mathbf{i} - f'_y(x_0, y_0)\mathbf{j}$ (см. п. 1). Через точку M_0 проводим луч в этом направлении, т. е. рассматриваем значение $f(x_0 - f'_{x0}t,\ y_0 - f'_{y0}t)$ как функцию от $t > 0$ и находим значение t, при котором эта функция одной переменной имеет минимум. Это значение определяет точку $M_1(x_1, y_1)$, через которую проводим луч в направлении $-(\operatorname{grad} f)_{M_1}$, ищем на нем точку условного минимума f и т. д. Во многих случаях после нескольких шагов мы получаем искомую точку безусловного минимума с хорошей точностью. (Продумайте геометрический смысл этого метода на плоскости x, y с нанесенными линиями уровня функции f.) Подобные методы нахождения экстремумов без обращения к необходимым условиям называются *прямыми методами.*

11. Экстремум с ограничениями. На независимые переменные могут быть наложены также одно или несколько соотношений, имеющих вид неравенств; такие соотношения называются *ограничениями* или *высвобождающими связями.* Пусть, например, ищется экстремум функции $f(x, y)$, причем независимые переменные связаны ограничением $F(x, y) \geqslant 0$, определяющим в плоскости x, y

некоторую область (S) с границей (L) (рис. 241), на которой $F \equiv 0$. Функция f может иметь как *внутренние* экстремумы, достигающиеся внутри (S), так и граничные экстремумы, достигающиеся на (L). Для нахождения первых можно воспользоваться условиями стационарности (19); однако для граничных экстремумов эти условия не действуют. Для нахождения граничных экстремумов можно заметить, что если в некоторой точке M (рис. 241) функция f имеет экстремум, например минимум, то значение $f(M)$ меньше всех значений f на (L) вблизи M. Поэтому в точке M одновременно достигается минимум f при условии $F=0$, а такие точки можно разыскать по методу п. 10.

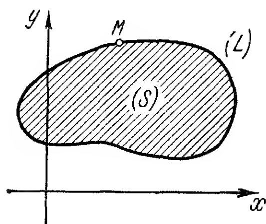


Рис. 241.

Полученные результаты можно применить *к нахождению наибольшего и наименьшего значений функции нескольких переменных*. Например, пусть функция $z=f(x, y)$ рассматривается в области, изображенной на рис. 242, причем предположим, что ни функция, ни ее производные не имеют в этой области разрывов. Если точка, в которой функция принимает наибольшее значение, находится внутри области, то в этой точке будет безусловный максимум функции. Если эта точка находится на контуре, но не в вершинах A, B, C, то в ней будет условный максимум, причем, за условие (32) нужно взять уравнение соответствующей дуги контура. Наконец, наибольшее значение может достигаться в вершинах A, B или C.

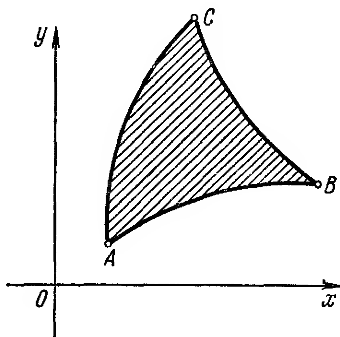


Рис. 242.

Таким образом, для отыскания этого значения надо найти все точки безусловного максимума внутри области и все точки условного максимума (или даже все точки возможного условного экстремума, если заранее трудно сказать, какие именно из этих точек будут точками максимума) на контуре и сравнить значения функции f во всех этих точках и в точках A, B и C. Наибольшее из этих значений (*наибольший максимум*) и даст наибольшее значение функции. Аналогично ищется наименьшее значение функции; как и в п. IV.19, удобно искать наибольшее и наименьшее значения одновременно.

Если область содержит точки разрыва производных первого порядка, то значение функции в таких точках должно быть «принято к сравнению», так как может оказаться, что наибольшее значение функции достигается именно там. Если имеются точки разрыва функции, то надо дополнительно исследовать предельные значения функции при приближении к таким точкам. Если имеются лин и разрыв фун ц л ее р изв дных первого порядка, то до исследо-

вать значения функции вдоль таких линий, что приводит к задаче об условном экстремуме. Наконец, если область, на которой рассматривается функция, простирается в бесконечность, то надо дополнительно исследовать предельные значения функции, когда точка в плоскости аргументов удаляется в бесконечность.

Подобным образом рассматриваются функции большего числа переменных. Но при этом надо иметь в виду, что если, например, функция $u=f(x, y, z)$ рассматривается в области, изображенной на рис. 243, то надо искать безусловные максимумы внутри «кривого тетраэдра», условные максимумы с одним условием на его «гранях» (при этом условием служит уравнение соответствующей грани) и условные максимумы с двумя условиями на «ребрах», при этом условиями служат уравнения ребер, если их записать в форме (X 2). Подобным образом поверхности разрыва приводят к исследованию условного максимума с одним условием, а линии разрыва — к условному максимуму с двумя условиями.

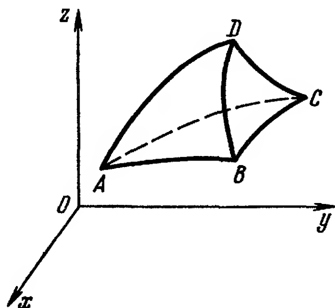


Рис. 243.

Если применяется итерационный метод типа метода наискорейшего спуска, описанного в п. 10, то при наличии нескольких максимумов мы можем прийти не к наибольшему. Здесь полезно провести метод несколько раз, начиная от различных, выбираемых наугад, нулевых приближений. После таких повторных вычислений и сравнения результатов возможно увеличить максимум, полученный при первом вычислении, а во многих задачах получить и наибольший максимум.

12. Численное решение систем уравнений. Рассмотрим в заключение некоторые методы численного решения системы двух уравнений с двумя неизвестными; случай системы n уравнений с n неизвестными рассматривается аналогично.

Метод итераций имеет тот же вид, что в пп. V.3 и VI.5. Для его применения заданная система записывается в виде

$$\left.\begin{aligned} x&=f(x, y),\\ y&=g(x, y). \end{aligned}\right\} \qquad (42)$$

Затем выбирается некоторое нулевое приближение $x=x_0$, $y=y_0$. Последующие приближения строятся по формулам

$$\left.\begin{aligned} x_1&=f(x_0, y_0);\\ y_1&=g(x_0, y_0); \end{aligned}\right\} \quad \left.\begin{aligned} x_2&=f(x_1, y_1);\\ y_2&=g(x_1, y_1) \end{aligned}\right\} \text{ и т. д.}$$

Если процесс сходится, то в пределе получится решение системы (42); сходимость процесса тем лучше, чем медленнее меняются функции f и g при изменении их аргументов, т. е. чем меньше абсолютная величина производных от этих функций.

Иногда некоторое ускорение сходимости удается получить, если на каждом итерационном шаге использовать результаты, полученные уже на этом шаге, т. е. вычислять так (это *метод Зайделя*): $x_1=f(x_0, y_0)$, $y_1=g(x_1, y_0)$, $x_2=f(x_1, y_1)$, $y_2=g(x_2, y_1)$ и т. д.

Метод Ньютона (см. п. V.2) состоит в замене заданных функций линейными на основе значений этих функций и их производных при значениях аргументов, приближенно равных искомым решениям. Пусть мы решаем систему

уравнений

$$\left.\begin{aligned} P(x,\ y)=0,\\ Q(x,\ y)=0,\end{aligned}\right\} \qquad (43)$$

отправляясь от некоторого нулевого приближения $x=x_0$, $y=y_0$ [его можно найти из физического смысла системы или из ориентировочного наброска линий (43) на плоскости x, y и т. п.] к искомому решению. Применяя разложение (17) функций P и Q по степеням $h=x-x_0$ и $k=y-y_0$ и отбрасывая члены высшего порядка малости, получим систему уравнений:

$$\left.\begin{aligned} P(x_0,\ y_0)+P'_x(x_0,\ y_0)(x-x_0)+P'_y(x_0,\ y_0)(y-y_0)=0,\\ Q(x_0,\ y_0)+Q'_x(x_0,\ y_0)(x-x_0)+Q'_y(x_0,\ y_0)(y-y_0)=0,\end{aligned}\right\} \qquad (44)$$

приближенно заменяющую систему (43). Решая систему (44), которая представляет собой систему *линейных* уравнений, получаем значения первого приближения $x=x_1$, $y=y_1$. Второе приближение находится из системы (44), если в ней x_0, y_0 заменить на x_1, y_1 и т. д. Связь между n-м и $(n+1)$-м приближениями имеет вид

$$\left.\begin{aligned} P(x_n,\ y_n)+P'_x(x_n,\ y_n)(x_{n+1}-x_n)+P'_y(x_n,\ y_n)(y_{n+1}-y_n)=0;\\ Q(x_n,\ y_n)+Q'_x(x_n,\ y_n)(x_{n+1}-x_n)+Q'_y(x_n,\ y_n)(y_{n+1}-y_n)=0.\end{aligned}\right\}$$

Если процесс сходится, то в пределе при $n\to\infty$ в каждом из этих равенств два последних слагаемых отпадают, т. е. получается решение системы (43). Можно проверить, что сходимость приближений имеет место, если нулевое приближение лежит в достаточной близости от искомого решения и якобиан (см. п. IX.13)

$$\frac{D(P,\ Q)}{D(x,\ y)}\neq 0.$$

[Чем является этот якобиан для системы (44)?]

Можно воспользоваться также тем, что решение системы (43) одновременно реализует минимум функции $V(x,\ y)=[P(x,\ y)]^2+[Q(x,\ y)]^2$ (почему?); впрочем, иногда перед квадратами ставятся положительные коэффициенты, чтобы в определенном смысле уравнять «значимость» обоих уравнений (43). После этого точка минимума функции V ищется по одному из прямых методов, о которых говорилось в пп. 10 и 11. (Конечно, применять необходимые условия (19) нельзя, так как мы придем опять к системе (43).) Если соответствующее минимальное значение равно нулю, то точка минимума и даст решение системы (43).

НЕОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ

§ 1. Элементарные методы интегрирования

1. Основные определения. Пусть функция $f(x)$ является производной от функции $F(x)$, т. е. $F'(x)=f(x)$; тогда $F(x)$ называется *первообразной функцией от* $f(x)$ (или *для* $f(x)$ и т. п.). Например, функция $3x^2$ является производной для x^3, а x^3 является первообразной для $3x^2$.

Основная задача дифференциального исчисления — это задача о нахождении производной и непосредственно связанная с ней задача о нахождении дифференциала заданной функции. Эта задача для функций одного переменного рассматривалась в гл. IV; в частности, в п. IV.5 мы видели, что производная от любой элементарной функции вычисляется по стандартным правилам и является элементарной функцией. *Основная задача интегрального исчисления — это задача о нахождении первообразной для заданной функции,* т. е. о нахождении функции по заданной ее производной; огромная важность этой задачи будет ясна из гл. XIV. Эта задача сложнее, чем задача дифференцирования. (Вообще «обратные» задачи обычно сложнее «прямых»: например, задача об извлечении корня сложнее задачи о возведении в степень.) В частности, мы увидим, что первообразная от любой элементарной функции (а мы в этой главе будем заниматься только такими функциями), хотя и всегда существует, но далеко не всегда является элементарной функцией.

Первообразная у заданной функции не одна; например, не только $(x^3)'=3x^2$, но и $(x^3+5)'=3x^2$. (И в других примерах решение «обратных» задач часто бывает неоднозначным.) Вообще, если функция $f(x)$ имеет первообразные $F_1(x)$ и $F_2(x)$, то $F_1'=f$, $F_2'=f$, т. е. $F_1'-F_2'=0$, $(F_1-F_2)'=0$, $F_1-F_2=\text{const}$ (см. п. IV.17), $F_1=F_2+\text{const}$. Таким образом, *любые две первообразные к одной и той же функции отличаются друг от друга на постоянное слагаемое*; чтобы получить все первообразные для данной функции, надо взять какую-нибудь одну и прибавить к ней *произвольную постоянную*. Например, совокупность всех первообразных к функции $3x^2$ дается формулой x^3+C, где C — произвольная постоянная, придавая которой численные значения, мы получим индивидуальные первообразные: x^3, x^3+5, $x^3-\sqrt{2}$, $x^3+\frac{5}{6}$ и т. п.

Совокупность всех первообразных к функции $f(x)$ называется *неопределенным интегралом* от функции $f(x)$ и обозначается $\int f(x)\,dx$; смысл этого обозначения будет подробно раскрыт в п. XIV.2. Здесь $\int$ — *знак интеграла*, $f(x)$ — *подынтегральная функция*, $f(x)\,dx$ — *подынтегральное выражение*. Таким образом,

$$\text{если } F'(x)=f(x), \text{ то } \int f(x)\,dx=F(x)+C, \text{ и наоборот.} \quad (1)$$

Например, $\int 3x^2dx=x^3+C$. Другими словами, неопределенный интеграл — это *общая первообразная*, содержащая произвольную постоянную, при каждом численном значении которой получается *частная первообразная*.

Из формулы (1) следует, что

$$\left(\int f(x)dx\right)'=f(x); \quad d\left(\int f(x)dx\right)=f(x)\,dx;$$
$$\int (dF(x))=F(x)+C. \quad (2)$$

Таким образом, *знаки дифференциала и интеграла уничтожают друг друга. Результат вычисления неопределенного интеграла всегда можно проверить, взяв производную от ответа*; при этом должна получиться подынтегральная функция. Каждой формуле дифференциального исчисления (пп. IV.4—5) отвечает некоторая формула интегрального исчисления.

2. Простейшие интегралы. Простейшие интегралы получаются в результате обращения формул для производных основных элементарных функций (п. IV.5). Например, из формулы $(\sin x)'=\cos x$ получаем

$$\int \cos x\,dx=\sin x+C \quad (3)$$

(см. формулу (1)). Формулу $(\cos x)'=-\sin x$ лучше переписать в виде $(-\cos x)'=\sin x$, откуда получаем

$$\int \sin x\,dx=-\cos x+C.$$

Аналогично

$$\int \frac{1}{\cos^2 x}dx=\operatorname{tg} x+C; \text{ пишут просто } \int \frac{dx}{\cos^2 x}=\operatorname{tg} x+C,$$
$$\int \frac{dx}{\sin^2 x}=-\operatorname{ctg} x+C; \quad \int \frac{dx}{\sqrt{1-x^2}}=\arcsin x+C. \quad (4)$$

Из формулы $(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}$ вытекает, что

$$\int \frac{dx}{\sqrt{1-x^2}} = -\arccos x + C. \tag{5}$$

На первый взгляд кажется, что эта формула противоречит предыдущей. Но это не так: на основании формулы $\arcsin x + \arccos x = \frac{\pi}{2}$ (IV.18) из (4) следует, что

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C = -\arccos x + \frac{\pi}{2} + C =$$
$$= -\arccos x + C_1, \text{ где } C_1 = \frac{\pi}{2} + C.$$

Итак, дело в том, что в *правых частях формул* (4) *и* (5) *произвольные постоянные различны*. Такое различие формы ответов бывает и в других примерах неопределенных интегралов. Естественно, что из двух формул (4) и (5) надо пользоваться только одной, например (4).

Из дальнейших формул дифференцирования получаем

$$\int \frac{dx}{1+x^2} = \operatorname{arctg} x + C; \qquad \int \frac{dx}{x} = \ln x + C.$$

Недостатком последней формулы является то, что функция $\frac{1}{x}$, от которой берется первообразная, существует как при $x > 0$, так и при $x < 0$, тогда как правая часть существует только при $x > 0$. Однако легко проверить формулу дифференцирования $(\ln|x|)' = \frac{1}{x}$; действительно, при $x > 0$ будет $|x| = x$, и потому получаем обычную производную логарифма, а при $x < 0$ будет $|x| = -x$, т. е. $(\ln|x|)' = (\ln(-x))' = \frac{1}{-x}(-1) = \frac{1}{x}$. Поэтому как при $x > 0$, так и при $x < 0$ можно написать

$$\int \frac{dx}{x} = \ln|x| + C. \tag{6}$$

Далее получаем

$$\int a^x dx = \frac{a^x}{\ln a} + C; \quad \text{в частности,} \quad \int e^x dx = e^x + C;$$
$$\int x^{n-1} dx = \frac{x^n}{n} + C, \quad \text{т. е.} \quad \int x^n dx = \frac{x^{n+1}}{n+1} + C.$$

Конечно, последняя формула при $n = -1$ не годится, так как тогда знаменатель обращается в нуль. Но в этом случае интеграл приобретает вид $\int \frac{dx}{x}$ и вычисляется по формуле (6).

Далее,

$$\int \operatorname{ch} x\,dx = \operatorname{sh} x + C; \quad \int \operatorname{sh} x\,dx = \operatorname{ch} x + C; \quad \int \frac{dx}{\operatorname{ch}^2 x} = \operatorname{th} x + C;$$

$$\int \frac{dx}{\sqrt{x^2+1}} = \operatorname{arsh} x + C = \ln\left(x + \sqrt{x^2+1}\right) + C$$

(формула для $\operatorname{arsh} x$ выведена в п. 1.28).

Найденную сейчас формулу

$$\int \frac{dx}{\sqrt{x^2+1}} = \ln\left(x + \sqrt{x^2+1}\right) + C$$

можно доказать без всяких гиперболических функций, при помощи дифференцирования ответа. Более того, так как $(\ln|u|)' = \frac{1}{u} u'_x$, то

$$\left(\ln\left|x+\sqrt{x^2+a}\right|\right)' = \frac{1}{x+\sqrt{x^2+a}}\left(1 + \frac{2x}{2\sqrt{x^2+a}}\right) = \frac{1}{\sqrt{x^2+a}},$$

откуда

$$\int \frac{dx}{\sqrt{x^2+a}} = \ln\left|x+\sqrt{x^2+a}\right| + C$$

при постоянном a любого знака.

Выведенные формулы (*табличные интегралы*) следует выписать в виде таблицы и запомнить наизусть, так как они широко применяются при вычислении интегралов. В частности, с помощью этих формул вычисляются (или, как говорят, *берутся*) некоторые «почти табличные» интегралы. Для этого исходят из табличной формулы и стараются изменить ответ так, чтобы производная от него давала подынтегральную функцию; этот метод *непосредственного интегрирования*, по существу, состоит просто в применении формулы (1). Например, при вычислении интеграла

$$\int \cos 3x\,dx \tag{7}$$

естественно вспомнить формулу (3). Однако ответ $\int \cos 3x\,dx = \sin 3x + C$ неправильный, так как производная от правой части равна $3\cos 3x$, а не $\cos 3x$, как должно быть. Но если правую часть разделить на 3, то и производная разделится на 3. Итак, $\left(\frac{1}{3}\sin 3x\right)' = \cos 3x$, т. е. $\int \cos 3x\,dx = \frac{1}{3}\sin 3x + C$.

Подобным образом найдем (проверьте ответ дифференцированием!)

$$\int \frac{dx}{\sqrt{1-(2x+5)^2}} = \frac{1}{2}\arcsin(2x+5) + C, \quad \int \frac{dx}{x-3} = \ln|x-3| + C,$$

$$\int \frac{dx}{1+\frac{x^2}{2}} = \int \frac{dx}{1+\left(\frac{x}{\sqrt{2}}\right)^2} = \sqrt{2}\operatorname{arctg}\frac{x}{\sqrt{2}} + C \text{ и т. д.} \tag{8}$$

Вообще, если найден интеграл $\int f(x)\,dx=F(x)+C$, то

$$\int f(ax+b)\,dx=\frac{1}{a}F(ax+b)+C \qquad (a,\ b=\mathrm{const}).$$

3. Простейшие свойства неопределенного интеграла. Эти свойства вытекают из аналогичных свойств производной (п. IV.4). Например,

$$\int [f(x)\pm\varphi(x)]\,dx=\int f(x)\,dx\pm\int\varphi(x)\,dx, \tag{9}$$

т. е. *интеграл от суммы равен сумме интегралов.* Для доказательства надо взять производные от левой и от правой частей и заметить, что в силу первой формулы (2) и свойства «производная суммы равна сумме производных» результаты дифференцирования равны. Но если производные равны, то функции могут различаться лишь на постоянное слагаемое, которое в формуле (9) писать не нужно, так как знаки неопределенных интегралов включают в себя произвольные постоянные слагаемые.

Аналогично проверяется, что

$$\int Af(x)\,dx=A\int f(x)\,dx \qquad (A=\mathrm{const}), \tag{10}$$

т. е. *постоянный множитель можно выносить за знак интеграла.*

Применение формул (9) и (10) часто дает возможность представить заданный интеграл в виде суммы табличных интегралов, после чего произвести почленное интегрирование и написать общий ответ (это — так называемый *метод разложения*). Приведем несколько примеров:

$$\int(3x^3-2x+5)\,dx=\int(3x^3)\,dx-\int(2x)\,dx+\int 5\,dx=$$
$$=3\int x^3dx-2\int x\,dx+5\int dx=3\frac{x^4}{4}-2\frac{x^2}{2}+5x+C=$$
$$=\frac{3}{4}x^4-x^2+5x+C \tag{11}$$

(конечно, надо писать только одну постоянную, так как сумма произвольных постоянных дает произвольную постоянную);

$$\int\frac{dx}{a^2+x^2}=\frac{1}{a^2}\int\frac{dx}{1+\frac{x^2}{a^2}}=\frac{1}{a^2}\int\frac{dx}{1+\left(\frac{x}{a}\right)^2}=$$
$$=\frac{1}{a^2}a\,\mathrm{arctg}\,\frac{x}{a}+C=\frac{1}{a}\mathrm{arctg}\,\frac{x}{a}+C$$

(ср. пример (8)); аналогично (проверьте!)

$$\int\frac{dx}{\sqrt{a^2-x^2}}=\arcsin\frac{x}{a}+C \qquad (a>0). \tag{12}$$

Другие примеры:

$$\int \operatorname{tg}^2 x dx = \int \frac{\sin^2 x}{\cos^2 x} dx = \int \frac{1-\cos^2 x}{\cos^2 x} dx =$$
$$= \int \left(\frac{1}{\cos^2 x} - 1\right) dx = \int \frac{dx}{\cos^2 x} - \int dx = \operatorname{tg} x - x + C;$$

$$\int \frac{1}{x(x-1)} dx = \int \frac{x-(x-1)}{x(x-1)} dx = \int \left(\frac{1}{x-1} - \frac{1}{x}\right) dx =$$
$$= \ln|x-1| - \ln|x| + C = \ln\left|\frac{x-1}{x}\right| + C;$$

$$\int \frac{1}{x^2-a^2} dx = \int \frac{(x+a)-(x-a)}{2a(x-a)(x+a)} dx =$$
$$= \frac{1}{2a} \int \left(\frac{1}{x-a} - \frac{1}{x+a}\right) dx = \frac{1}{2a} \ln\left|\frac{x-a}{x+a}\right| + C.$$

Прием представления заданной дроби в виде суммы дробей более простого вида, примененный при вычислении двух последних интегралов, является довольно общим. Он состоит в том, что знаменатель разлагают на множители, после чего стараются представить числитель в виде комбинации множителей, стоящих в знаменателе. Если это удастся, то после разложения дроби на сумму нескольких дробей можно произвести сокращение в каждом из слагаемых.

Приведем еще полезный пример. Пусть надо вычислить $\int \sin 5x \cos 3x \, dx$. Из тригонометрии известна формула

$$\sin\alpha \cos\beta = \frac{1}{2}[\sin(\alpha+\beta) + \sin(\alpha-\beta)].$$

Поэтому

$$\int \sin 5x \cos 3x \, dx = \int \frac{\sin 8x + \sin 2x}{2} dx = -\frac{1}{16}\cos 8x - \frac{1}{4}\cos 2x + C.$$

В аналогичных случаях применяются также формулы

$$\sin\alpha \sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)];$$
$$\cos\alpha \cos\beta = \frac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)];$$
$$\sin^2\alpha = \frac{1-\cos 2\alpha}{2}; \qquad \cos^2\alpha = \frac{1+\cos 2\alpha}{2}.$$

Например,

$$\int \sin^2 3x \, dx = \int \frac{1-\cos 6x}{2} dx = \frac{1}{2}x - \frac{1}{12}\sin 6x + C.$$

Отметим еще один интересный прием, основанный на применении комплексных функций от вещественного аргумента (п. VIII.6), для которых все формулы интегрирования остаются в силе. Ясно, что если такую функцию проинтегрировать, то ее вещественная и мнимая части тоже проинтегриру-

ются, т. е. $\int \mathrm{Re} = \mathrm{Re} \int$, $\int \mathrm{Im} = \mathrm{Im} \int$. Это дает возможность, например, вычислить с помощью формулы Эйлера (п. VIII.4) такой вещественный интеграл:

$$\int e^{ax} \cos bx\, dx = \int \mathrm{Re}\, [e^{ax} e^{ibx}]\, dx = \mathrm{Re} \int e^{(a+ib)x}\, dx =$$
$$= \mathrm{Re} \frac{e^{(a+ib)x}}{a+ib} + C = \mathrm{Re} \frac{e^{ax} (\cos bx + i \sin bx)(a - ib)}{a^2+b^2} + C =$$
$$= e^{ax} \frac{a \cos bx + b \sin bx}{a^2+b^2} + C.$$

4. Интегрирование по частям. К сожалению, не существует формулы, выражающей интеграл от произведения функций через интегралы от сомножителей. С этим связано то обстоятельство, что в отличие от производных *интеграл от элементарной функции* (понятие элементарной функции см. в п. I.18) *не всегда является элементарной функцией*. Например, интегралы $\int \sin x\, dx$ и $\int \frac{1}{x}\, dx$ — табличные, тогда как интеграл $\int \frac{\sin x}{x}\, dx$ не выражается через основные элементарные функции («не берется в элементарных функциях»). Как быть с такими, неэлементарными функциями, мы скажем в п. 11.

Тем не менее, если проинтегрировать обе части формулы $(uv)' = u'v + uv'$ (п. IV.4), получится

$$uv = \int u'v\, dx + \int uv'\, dx,$$

т. е.

$$\int uv'\, dx = uv - \int u'v\, dx, \qquad (13)$$

или, что то же,

$$\int u\, dv = uv - \int v\, du. \qquad (14)$$

Формула (13) и равносильная ей формула (14) называются *формулами интегрирования по частям*. При применении формулы (13) подынтегральная функция разлагается на два множителя (u и v'), из которых один дифференцируется, а второй интегрируется; другими словами, мы переходим к интегралу, у которого вместо u стоит u', а вместо v' стоит v. После такого преобразования иногда может получиться табличный интеграл или интеграл более простой, чем исходный.

Приведем несколько примеров. При вычислении интеграла $\int x^2 \ln x\, dx$ выгодно продифференцировать $\ln x$, так как тогда получится степенная функция, которая проще логарифмической; правда, при этом второй множитель (x^2) придется интегрировать, но он и после интегрирования останется степенной функцией. Итак, обозначаем $u = \ln x$, $dv = x^2 dx$, откуда $v = \frac{x^3}{3}$ и

$$\int x^2 \ln x\, dx = \ln x \frac{x^3}{3} - \int \frac{x^3}{3}\, d(\ln x) = \frac{x^3}{3} \ln x - \int \frac{x^2}{3}\, dx = \frac{x^3}{3} \ln x - \frac{x^3}{9} + C.$$

Отметим, что при вычислении v не надо было писать произвольную постоянную, т. е. писать $v=\frac{x^3}{3}+C$, так как в формуле (13) v является какой-то одной, индивидуальной функцией.

Аналогичным образом часто стараются продифференцировать функции $\operatorname{arctg} x$ и $\arcsin x$, так как после этого получаются более простые функции.

При вычислении интеграла $\int x^2 \sin 3x\,dx$ следует дифференцировать степень, так как при этом показатель степени понижается на единицу (поэтому интегрировать по частям придется два раза); в то же время при дифференцировании синуса, как и при его интегрировании, он не упрощается и не усложняется:

$$\int x^2 \sin 3x\,dx = \left| \begin{matrix} u=x^2, & dv=\sin 3x dx \\ du=2x dx, & v=-\frac{1}{3}\cos 3x \end{matrix} \right| =$$

$$= -\frac{x^2}{3}\cos 3x + \int \frac{1}{3}\cos 3x \cdot 2x dx = \left| \begin{matrix} u=x, & dv=\cos 3x dx \\ du=dx, & v=\frac{1}{3}\sin 3x \end{matrix} \right| =$$

$$= -\frac{x^2}{3}\cos 3x + \frac{2}{3}\left(\frac{x}{3}\sin 3x - \int \frac{1}{3}\sin 3x dx\right) =$$

$$= -\frac{1}{3}x^2\cos 3x + \frac{2}{9}x\sin 3x + \frac{2}{27}\cos 3x + C.$$

Вспомогательные записи здесь отделены вертикальными черточками.

Бывает так, что после интегрирования по частям и преобразований в правой части получается исходный интеграл, но с другим коэффициентом. Тогда, приводя подобные члены, можно этот интеграл вычислить. Например,

$$\int \sqrt{1-x^2}\,dx = \left| \begin{matrix} u=\sqrt{1-x^2}, & dv=dx \\ du=-\frac{x}{\sqrt{1-x^2}}dx, & v=x \end{matrix} \right| =$$

$$= x\sqrt{1-x^2} + \int \frac{x^2}{\sqrt{1-x^2}}dx = x\sqrt{1-x^2} + \int \frac{(x^2-1)+1}{\sqrt{1-x^2}}dx =$$

$$= x\sqrt{1-x^2} - \int \sqrt{1-x^2}\,dx + \int \frac{1}{\sqrt{1-x^2}}dx =$$

$$= x\sqrt{1-x^2} - \int \sqrt{1-x^2}\,dx + \arcsin x.$$

Перенеся полученный интеграл в левую часть (отчего в правой части может остаться постоянное слагаемое, так как неопределенный интеграл известен лишь с точностью до такого слагаемого), получим

$$2\int \sqrt{1-x^2}\,dx = x\sqrt{1-x^2} + \arcsin x + C,$$

$$\int \sqrt{1-x^2}\,dx = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\arcsin x + C_1 \quad \left(C_1=\frac{C}{2}\right).$$

5. Замена переменной. Мы сейчас опишем один из наиболее распространенных приемов интегрального исчисления, основанный на формуле производной сложной функции (п. IV.4). Допустим, что

функция $F(x)$ является первообразной к $f(x)$, а x как-то зависит от переменной t, $x=\varphi(t)$. Найдем производную от $F(x)$ по t:

$$[F(x)]'_t=[F(x)]'_x\cdot x'_t=f(x)\,\varphi'(t)=f[\varphi(t)]\,\varphi'(t).$$

Если проинтегрировать обе части по t, то получим

$$F(x)+C=\int f[\varphi(t)]\,\varphi'(t)\,dt;$$

другими словами, в силу формул (1)

$$\left[\int f(x)\,dx\right]\Big|_{x=\varphi(t)}=\int f[\varphi(t)]\,\varphi'(t)\,dt. \tag{15}$$

Это и есть основная *формула замены переменной.*

Так как $\varphi'(t)\,dt=dx$, то правую часть формулы (15) можно переписать в виде $\int f(x)\,dx$; однако здесь в процессе интегрирования x не считается независимой переменной, а зависит от t. Поэтому формулу (15) можно истолковать так: *любая формула интегрирования вида*

$$\int f(x)\,dx=F(x)+C \tag{16}$$

сохраняет силу, если как в подынтегральном выражении, так и в правой части сделать произвольную замену переменной $x=\varphi(t)$. В этом смысле всякая формула вида (16) инвариантна.

Например, из формулы (3) после подстановки $x=u^3$ вытекает

$$\int \cos u^3\,d(u^3)=\sin u^3+C,\ \text{ т. е. }\ \int u^2\cos u^3\,du=\frac{1}{3}\sin u^3+C$$

и т. п. Однако при практическом применении формулы (15), конечно, не отправляются от табличной формулы, а, наоборот, стараются сделать такую подстановку, чтобы из заданного интеграла получился табличный.

Рассмотрим несколько примеров. В интеграле $\int\frac{\sqrt{x}}{1+x}\,dx$, чтобы избавиться от радикала, делаем замену переменной

$$\int\frac{\sqrt{x}}{1+x}\,dx=\left|\begin{matrix}x=t^2\\ dx=2t\,dt\end{matrix}\right|=\int\frac{t}{1+t^2}\cdot 2t\,dt=$$
$$=2\int\frac{t^2+1-1}{1+t^2}\,dt=2\left(\int dt-\int\frac{dt}{1+t^2}\right)=$$
$$=2(t-\operatorname{arctg}t+C)=2(\sqrt{x}-\operatorname{arctg}\sqrt{x})+C,$$

таким образом, после замены переменной и интегрирования надо выполнить обратную замену, т. е. перейти от t к x.

Формулу (15) часто читают «справа налево», т. е. делают замену не вида $x=\varphi(t)$, а вида $\psi(x)=u$. Например, чтобы вычислить интеграл $\int xe^{x^2}\,dx$, замечают, что подынтегральное выражение просто выражается через x^2, так как $x\,dx=\frac{1}{2}d(x^2)$; поэтому

$$\int xe^{x^2}dx=\left|\begin{matrix}x^2=u\\ 2x\,dx=du\end{matrix}\right|=\int e^u\frac{1}{2}du=\frac{1}{2}e^u+C=\frac{1}{2}e^{x^2}+C.$$

Применяются также замены вида $\psi(x)=\varphi(t)$.

Интеграл (7) можно было бы вычислить путем замены переменной

$$\int\cos 3x\,dx=\left|\begin{matrix}3x=t\\ 3dx=dt\end{matrix}\right|=\int\cos t\,\frac{dt}{3}=$$
$$=\frac{1}{3}\int\cos t\,dt=\frac{1}{3}\sin t+C=\frac{1}{3}\sin 3x+C.$$

Вычисление можно проводить, не выписывая замену явно:

$$\int\cos 3x\,dx=\int\cos 3x\frac{d(3x)}{3}=\frac{1}{3}\int\cos 3x\,d(3x)=\frac{1}{3}\sin 3x+C;$$

здесь мы непосредственно использовали инвариантность формулы (3).

Подобным образом,

$$\int\operatorname{tg}x\,dx=\int\frac{\sin x}{\cos x}dx=-\int\frac{d(\cos x)}{\cos x}=-\ln|\cos x|+C;$$

вообще

$$\int\frac{f'(x)}{f(x)}dx=\int\frac{df(x)}{f(x)}=\ln|f(x)|+C. \tag{17}$$

Далее,

$$\int\frac{x}{\sqrt{x^2+1}}dx=\int(x^2+1)^{-\frac{1}{2}}\cdot\frac{1}{2}d(x^2+1)=$$
$$=\frac{1}{2}\cdot\frac{(x^2+1)^{1/2}}{\frac{1}{2}}+C=\sqrt{x^2+1}+C;$$

вообще

$$\int\frac{f'(x)}{\sqrt{f(x)}}dx=\int[f(x)]^{-1/2}df(x)=\frac{[f(x)]^{1/2}}{\frac{1}{2}}+C=2\sqrt{f(x)}+C. \tag{18}$$

С помощью формул (17) и (18) и дополнения до полного квадрата вычисляются, в частности, часто встречающиеся интегралы

$$\int\frac{ax+b}{\sqrt{px^2+qx+r}}dx \quad \text{и} \quad \int\frac{ax+b}{px^2+qx+r}dx.$$

Покажем, например, вычисление интеграла $\int\frac{2x-3}{\sqrt{-3x^2+2x+1}}dx$; для

этого заметим, что производная подкоренного выражения равна $-6x+2=$ $=-6\left(x-\frac{1}{3}\right)$:

$$\int \frac{2x-3}{\sqrt{-3x^2+2x+1}}\,dx=\int \frac{2\left[\left(x-\frac{1}{3}\right)+\frac{1}{3}\right]-3}{\sqrt{-3x^2+2x+1}}\,dx=$$

$$=\int \frac{2\left(x-\frac{1}{3}\right)}{\sqrt{-3x^2+2x+1}}\,dx+\int \frac{-\frac{7}{3}}{\sqrt{-3x^2+2x+1}}\,dx=$$

$$=-\frac{1}{3}\int \frac{-6\left(x-\frac{1}{3}\right)}{\sqrt{-3x^2+2x+1}}\,dx-\frac{7}{3\sqrt{3}}\int \frac{dx}{\sqrt{-\left[x^2-\frac{2}{3}x-\frac{1}{3}\right]}}=$$

$$=-\frac{2}{3}\sqrt{-3x^2+2x+1}-\frac{7}{3\sqrt{3}}\int \frac{d\left(x-\frac{1}{3}\right)}{\sqrt{\frac{4}{9}-\left(x-\frac{1}{3}\right)^2}}=$$

$$=-\frac{2}{3}\sqrt{-3x^2+2x+1}-\frac{7}{3\sqrt{3}}\arcsin\frac{x-\frac{1}{3}}{\frac{2}{3}}+C=$$

$$=-\frac{2}{3}\sqrt{-3x^2+2x+1}-\frac{7}{3\sqrt{3}}\arcsin\frac{3x-1}{2}+C \qquad (19)$$

(см. формулу (12)).

Вычисление интегралов — более сложное дело, чем вычисление производных; чтобы овладеть элементарными методами интегрирования, требуется проделать большое число упражнений.

§ 2. Систематическое интегрирование

В этом параграфе будет указано несколько классов функций, интегралы от которых можно вычислять стандартными методами. Следует иметь в виду, что эти стандартные методы не всегда являются самыми простыми; во многих случаях те или иные предварительные преобразования или непосредственное применение методов § 1 существенно упрощают вычисления. Сознательный выбор простейшего пути вычисления интегралов можно осуществить только после необходимой практики в интегрировании.

6. Интегрирование рациональных функций. Интегрирование рациональных функций производится на основе результатов п. VIII.10. Как мы видели, всякую дробно-рациональную функцию можно представить в виде суммы целой части, если дробь неправильная, и простейших рациональных дробей первого и второго типов. Целая часть, т. е. многочлен, интегрируется почленно простейшим способом (см. пример (11)), а интегрирование простейших дробей первого типа, которые имеют общий вид $A/(x-a)^{\alpha}$, немногим сложнее.

Например, если нужно проинтегрировать функцию (VIII.38), то в силу (VIII.39) и (VIII.42) получим

$$\int \frac{x^3-2x+3}{x(x-1)(x+2)^2}\,dx=\int\left[-\frac{3}{4}\,\frac{1}{x}+\frac{2}{9}\,\frac{1}{(x-1)}-\frac{1}{6}\,\frac{1}{(x+2)^2}+\frac{55}{36}\,\frac{1}{(x+2)}\right]dx=$$
$$=-\frac{3}{4}\ln|x|+\frac{2}{9}\ln|x-1|+\frac{1}{6}\,\frac{1}{(x+2)}+\frac{55}{36}\ln|x+2|+\text{const}.$$

Таким образом, остаются простейшие дроби второго типа, общий вид которых таков:

$$\frac{Mx+N}{(x^2+px+q)^\beta}. \tag{20}$$

Интегрирование такой дроби начинают с упрощения числителя. Для этого, имея в виду, что $(x^2+px+q)'=2\left(x+\frac{p}{2}\right)$, заменяют в числителе x на $\left(x+\frac{p}{2}\right)-\frac{p}{2}$, после чего приводят подобные члены, не раскрывая круглых скобок. Если затем разбить интеграл на два (ср. аналогичные вычисления (19)), то первый имеет вид

$$\int\frac{\left(x+\frac{p}{2}\right)dx}{(x^2+px+q)^\beta}=\frac{1}{2}\int\frac{d\,(x^2+px+q)}{(x^2+px+q)^\beta}$$

и потому сразу берется. Во втором интеграле, который имеет вид $\int\frac{dx}{(x^2+px+q)^\beta}$, надо произвести в знаменателе дополнение до полного квадрата, что даст $x^2+px+q=(x+a)^2+b$, где a и b — некоторые постоянные. Если затем обозначить $x+a=y$, то получится интеграл

$$I_\beta=\int\frac{1}{(y^2+b)^\beta}\,dy, \tag{21}$$

который при $\beta=1$ легко берется (как?). Если же $\beta=2,\ 3,\ \ldots$, то, как мы сейчас выведем, с помощью интегрирования по частям возможно перейти от I_β к более простому интегралу $I_{\beta-1}$:

$$I_\beta=\frac{1}{b}\int\frac{b}{(y^2+b)^\beta}\,dy=\frac{1}{b}\int\frac{(b+y^2)-y^2}{(y^2+b)^\beta}\,dy=\frac{1}{b}\,I_{\beta-1}-\frac{1}{b}\int y\,\frac{y}{(y^2+b)^\beta}\,dy=$$
$$=\left|\begin{matrix} u=y, & dv=\dfrac{y}{(y^2+b)^\beta}\,dy, \\ du=dy, & v=\displaystyle\int\frac{y\,dy}{(y^2+b)^\beta}=\frac{1}{2}\int\frac{d\,(y^2+b)}{(y^2+b)^\beta}=\frac{-1}{2(\beta-1)}\,\frac{1}{(y^2+b)^{\beta-1}}\end{matrix}\right|=$$
$$=\frac{1}{b}\,I_{\beta-1}+\frac{y}{2b(\beta-1)(y^2+b)^{\beta-1}}-\frac{1}{b}\int\frac{1}{2(\beta-1)}\,\frac{1}{(y^2+b)^{\beta-1}}\,dy=$$
$$=\frac{y}{2b(\beta-1)(y^2+b)^{\beta-1}}+\frac{2\beta-3}{2b(\beta-1)}\,I_{\beta-1} \tag{22}$$

(проверьте эти вычисления!)

Формулы такого типа, *в которых неизвестная величина, имеющая номер* (в данном случае I_β имеет номер β), *выражается через аналогичные величины с меньшими номерами, называются рекурре́нтными формулами.* Они не дают сразу решение задачи, но часто позволяют получить это решение после нескольких шагов, при которых номер неизвестных последовательно уменьшается. Например, в формуле (22) I_β выражено через $I_{\beta-1}$. Если эту формулу применить к $I_{\beta-1}$, т. е. подставить $\beta-1$ вместо β, то мы выразим $I_{\beta-1}$ через $I_{\beta-2}$ и т. д., пока не дойдем до I_1, а этот интеграл, как сказано выше, сразу берется.

Интересно отметить, что в определении простейшей дроби второго типа требовалось, чтобы корни знаменателя были мнимыми, но в процессе вычисления интеграла это не было использовано. Поэтому тем же методом можно интегрировать дроби вида (20) при вещественных корнях знаменателя, без разложения ее на простейшие дроби первого типа.

Интеграл (21) при $b>0$ можно взять также с помощью подстановки $y=\sqrt{b}\operatorname{tg} t$; тогда получится интеграл от степени $\cos t$, о котором мы скажем позже.

Итак, *интеграл от рациональной функции всегда может быть, и притом стандартным способом, выражен через элементарные функции,* а именно, рациональные функции, логарифм и арктангенс. Практически наиболее трудным шагом в этом интегрировании является разложение знаменателя по формуле (VIII.29).

Метод вычисления многих других интегралов, к которым мы сейчас перейдем, заключается в том, что после выполнения той или иной подстановки мы переходим к интегралу от рациональной функции (как говорят, происходит *рационализация* заданного интеграла), после чего интеграл берется по описанному стандартному способу.

7. Линейные и дробно-линейные иррациональности. Рассмотрим интеграл вида

$$\int R\left(x, \sqrt[n]{ax+b}\right)dx \qquad (n=2,\ 3,\ \ldots), \tag{23}$$

где a и b — постоянные, а $R(x, y)$ — рациональная функция относительно своих двух аргументов (см. п. I. 17). Тогда вся подынтегральная функция, рассмотренная как сложная функция от x, будет из-за наличия радикала иррациональной. Для рационализации этого интеграла совершим подстановку

$$ax+b=t^n, \quad a\,dx=nt^{n-1}\,dt,$$

$$\int R\left(x, \sqrt[n]{ax+b}\right)dx=\int \underline{R\left(\frac{t^n-b}{a},\ t\right)nt^{n-1}}\,dt.$$

Полученная подчеркнутая подынтегральная функция уже является рациональной (почему?).

Аналогичным образом, интеграл

$$\int R\left(x, \sqrt[n]{ax+b},\quad \sqrt[m]{ax+b},\ \ldots\right)dx \qquad (n,\ m=2, 3, 4, \ldots), \tag{24}$$

где $R(x, y, z, \ldots)$ — рациональная функция от своих аргументов, переходит в интеграл от рациональной функции после подстановки $ax+b=t^p$ с соответственно подобранным p (какое нужно взять в общем случае p?).

Например,

$$\int \frac{dx}{\sqrt{2x+3}-2\sqrt[3]{2x+3}} = \left| \begin{matrix} 2x+3=t^6 \\ 2\,dx=6t^5\,dt \end{matrix} \right| = \int \frac{3t^5\,dt}{t^3-2t^2} = 3\int \frac{t^3}{t-2}\,dt =$$

$$= \left| \begin{array}{l} t^3 \quad \mid t-2 \\ t^3-2t^2 \mid t^2+2t+4 \\ \quad 2t^2 \\ \quad 2t^2-4t \\ \qquad 4t \\ \qquad 4t-8 \\ \qquad\quad 8 \end{array} \right| \begin{array}{l} =3\int \left(t^2+2t+4+\dfrac{8}{t-2}\right) dt = \\ = t^3+3t^2+12t+24\ln|t-2|+C = \\ = \sqrt{2x+3}+3\sqrt[3]{2x+3}+12\sqrt[6]{2x+3}+ \\ +24\ln\left|\sqrt[6]{2x+3}-2\right|+C. \end{array}$$

Рационализация интеграла

$$\int R\left(x, \sqrt[n]{\frac{ax+b}{cx+d}}\right) dx \quad (n=2,\ 3,\ \ldots), \tag{25}$$

где $R(x, y)$ — рациональная функция, осуществляется с помощью подстановки

$$\frac{ax+b}{cx+d}=t^n, \quad ax+b=cxt^n+d\cdot t^n, \quad x=\frac{d\cdot t^n-b}{a-ct^n}.$$

Итак, *интегралы* (23)—(25), *в которых* R — *знак рациональной функции от своих аргументов, всегда выражаются через элементарные функции.*

8. Квадратичные иррациональности. *Интеграл вида*

$$\int R\left(x, \sqrt{ax^2+bx+c}\right) dx, \tag{26}$$

где $R(x, y)$ — *рациональная функция относительно своих аргументов, также всегда выражается через элементарные функции.* При вычислении такого интеграла часто применяются тригонометрические подстановки. Для этого надо произвести дополнение до полного квадрата, в результате чего получится

$$\int R\left(x, \sqrt{ax^2+bx+c}\right) dx = \int R\left(x, \sqrt{\pm(kx+l)^2 \pm m^2}\right) dx,$$

где k, l, m — какие-то постоянные. Затем делается подстановка

в случае $\sqrt{(kx+l)^2+m^2}$ $\qquad kx+l=m\operatorname{tg} t$,

в случае $\sqrt{-(kx+l)^2+m^2}$ $\qquad kx+l=m\sin t$,

в случае $\sqrt{(kx+l)^2-m^2}$ $\qquad kx+l=\dfrac{m}{\cos t}$;

случая $\sqrt{-(kx+l)^2-m^2}$ для вещественных интегралов не может быть. После этого корни извлекаются (проверьте!) и получается интеграл вида

$$\int R_1(\cos t,\ \sin t)\,dt, \tag{27}$$

где $R_1(x,\ y)$ — некоторая новая рациональная функция относительно своих аргументов. Как брать интегралы вида (27), будет описано в п. 10.

Используются также гиперболические подстановки:

$$kx+l=m\,\mathrm{sh}\,t,\quad m\,\mathrm{th}\,t,\quad m\,\mathrm{ch}\,t.$$

При вычислении интегралов вида (26) применяются также более прямые методы. Так, нередко можно перейти к интегралу вида

$$\int \frac{P_n(x)}{\sqrt{ax^2+bx+c}}\,dx, \tag{28}$$

где $P_n(x)$ — многочлен степени n. Этот интеграл легко берется с помощью метода неопределенных коэффициентов. Покажем, что интеграл (28) можно представить в виде

$$Q_{n-1}(x)\sqrt{ax^2+bx+c}+K\int\frac{dx}{\sqrt{ax^2+bx+c}}, \tag{29}$$

где $Q_{n-1}(x)$ — некоторый многочлен степени $n-1$, а K — постоянная; последний же интеграл легко берется (см. конец п. 5).

Пусть, например, $n=3$; приравнивая (28) и (29), получим

$$\int\frac{\alpha x^3+\beta x^2+\gamma x+\delta}{\sqrt{ax^2+bx+c}}\,dx=(Ax^2+Bx+C)\sqrt{ax^2+bx+c}+ \\ +K\int\frac{dx}{\sqrt{ax^2+bx+c}}, \tag{30}$$

где все коэффициенты в левой части заданы, а коэффициенты A, B, C, K надо найти. Для этого продифференцируем равенство (30):

$$\frac{\alpha x^3+\beta x^2+\gamma x+\delta}{\sqrt{ax^2+bx+c}}= \\ =(2Ax+B)\sqrt{ax^2+bx+c}+(Ax^2+Bx+C)\frac{2ax+b}{2\sqrt{ax^2+bx+c}}+\frac{K}{\sqrt{ax^2+bx+c}},$$

$$\alpha x^3+\beta x^2+\gamma x+\delta=(2Ax+B)(ax^2+bx+c)+\frac{1}{2}(Ax^2+Bx+C)(2ax+b)+K.$$

Приравнивание коэффициентов при одинаковых степенях x даст:

$$\left.\begin{array}{l|l} x^3 & 3aA=\alpha, \\ x^2 & \frac{5}{2}bA+2aB=\beta, \\ x & 2cA+\frac{3}{2}bB+aC=\gamma, \\ 1 & cB+\frac{1}{2}bC+K=\delta. \end{array}\right\}$$

Так как $a \neq 0$, то из первого уравнения легко найти A; подставив найденное значение во второе уравнение, находим B и т. д. Так мы находим все коэффициенты A, B, C, K и тем самым обосновываем формулу (30).

Пусть, например, надо вычислить интеграл

$$I = \int \sqrt{2x^2 - 2x + 1}\, dx.$$

Для этого пишем

$$\int \sqrt{2x^2-2x+1}\, dx = \int \frac{2x^2-2x+1}{\sqrt{2x^2-2x+1}}\, dx = (Ax+B)\sqrt{2x^2-2x+1} + K\int \frac{dx}{\sqrt{2x^2-2x+1}};$$

$$\frac{2x^2-2x+1}{\sqrt{2x^2-2x+1}} = A\sqrt{2x^2-2x+1} + (Ax+B)\frac{4x-2}{2\sqrt{2x^2-2x+1}} + \frac{K}{\sqrt{2x^2-2x+1}};$$

$$2x^2-2x+1 = A(2x^2-2x+1) + (Ax+B)(2x-1) + K,$$

$$\begin{array}{c|l} x^2 & 4A = 2, \\ x & -3A + 2B = -2, \\ 1 & A - B + K = 1, \end{array}\Bigg\}, \text{ откуда } A = \frac{1}{2},\quad B = -\frac{1}{4},\quad K = \frac{1}{4};$$

$$\int \frac{dx}{\sqrt{2x^2-2x+1}} = \frac{1}{\sqrt{2}}\int \frac{dx}{\sqrt{x^2 - x + \frac{1}{2}}} =$$

$$= \frac{1}{\sqrt{2}}\int \frac{d\left(x-\frac{1}{2}\right)}{\sqrt{\left(x-\frac{1}{2}\right)^2 + \frac{1}{4}}} = \frac{1}{\sqrt{2}}\ln\left|x - \frac{1}{2} + \sqrt{\left(x-\frac{1}{2}\right)^2 + \frac{1}{4}}\right| + C =$$

$$= \frac{1}{\sqrt{2}}\ln\frac{|2x-1+\sqrt{2x^2-2x+1}|}{2} + C =$$

$$= \frac{1}{\sqrt{2}}\ln|2x-1+\sqrt{2x^2-2x+1}| + C_1 \quad \left(C_1 = -\frac{1}{\sqrt{2}}\ln 2 + C\right).$$

Окончательно

$$I = \left(\frac{1}{2}x - \frac{1}{4}\right)\sqrt{2x^2-2x+1} + \frac{1}{4\sqrt{2}}\ln|2x-1+\sqrt{2x^2-2x+1}| + C_2 \quad \left(C_2 = \frac{1}{4}C_1\right).$$

Конечно, в ответе вместо C_2 можно писать просто C.

Интеграл вида

$$\int \frac{dx}{(x-\alpha)^n\sqrt{ax^2+bx+c}} \quad (n = 1,\ 2,\ \ldots) \tag{31}$$

приводится к интегралу (28) после подстановки $x - \alpha = \frac{1}{t}$. Значит, и здесь после этой подстановки можно применить метод неопределенных коэффициентов. Его можно применить и к интегралу вида

$$\int \frac{Q(x)}{P(x)}\frac{1}{\sqrt{ax^2+bx+c}}\, dx, \tag{32}$$

где $P(x)$ и $Q(x)$ — некоторые многочлены, если разложить дробь $Q(x)/P(x)$ на целую часть и сумму простейших рациональных дробей первого типа (см. формулу (VIII. 37)). Тогда весь интеграл (32) разобьется на сумму интегралов вида (28) и (31).

9. **Дифференциальный бином.** Здесь будет идти речь об интегралах вида

$$I=\int (ax^n+b)^p x^m\,dx,$$

где все показатели n, p, m — числа рациональные, т. е. целые или дробные. Эйлер в 1768 г. указал *три случая интегрируемости*:

1. *Число p целое.* Если $p>0$, то надо просто раскрыть скобку и произвести почленное интегрирование. Если же $p<0$, то надо совершить подстановку $x=t^k$, причем целое число k подобрать таким, чтобы все показатели стали целыми, т. е. чтобы получился интеграл от рациональной функции.

2. *Число $\frac{m+1}{n}$ целое.* Совершим подстановку $ax^n+b=u$; тогда

$$I=\int u^p\left[\left(\frac{u-b}{a}\right)^{1/n}\right]^m\cdot\frac{1}{n}\left(\frac{u-b}{a}\right)^{\frac{1}{n}-1}\cdot\frac{1}{a}\,du=$$

$$=\frac{1}{na^{\frac{m+1}{n}}}\int (u-b)^{\frac{m+1}{n}-1}u^p\,du,$$

и так как показатель степени у полученной скобки — число целое, то мы приходим к предыдущему случаю 1.

3. *Число $\frac{m+1}{n}+p$ целое.* Здесь надо совершить подстановку $ax^n+b=ux^n$, после чего мы вновь приходим к случаю 1; это мы предоставим сделать читателю.

Лишь в 1853 г. П. Л. Чебышев доказал, что, за исключением этих трех случаев, интеграл от дифференциального бинома не выражается через элементарные функции (см. п. 11).

10. Интегрирование функций, рационально зависящих от тригонометрических функций. *Интеграл вида*

$$\int R(\sin x,\ \cos x)\,dx, \tag{33}$$

где $R(u,\ v)$ — рациональная функция от u, v, всегда выражается через элементарные функции. Чтобы убедиться в этом, сделаем подстановку $\operatorname{tg}\frac{x}{2}=t$; тогда (проверьте!)

$$\sin x=\frac{2\operatorname{tg}\frac{x}{2}}{1+\operatorname{tg}^2\frac{x}{2}}=\frac{2t}{1+t^2},\quad \cos x=\frac{1-t^2}{1+t^2},\quad dx=\frac{2dt}{1+t^2}, \tag{34}$$

т. е. из (33) получится интеграл

$$\int R\left(\frac{2t}{1+t^2},\ \frac{1-t^2}{1+t^2}\right)\frac{2}{1+t^2}\,dt$$

от рациональной функции t, который можно взять по методу п. 6.

Универсальная подстановка (34) часто приводит к довольно громоздким рациональным дробям и потому на практике применяется не очень часто. В некоторых, менее общих случаях лучше применять другие подстановки, о которых мы сейчас скажем.

1. Пусть подынтегральная функция в (33) нечетна относительно $\sin x$, т. е. $R(-\sin x, \cos x) \equiv -R(\sin x, \cos x)$. Если теперь написать

$$I = \int R(\sin x, \cos x)\, dx = \int \frac{R(\sin x, \cos x)}{\sin x} \sin x\, dx = \\ = \int R_1(\sin x, \cos x) \sin x\, dx,$$

то функция R_1 будет уже четной относительно $\sin x$ и ее легко выразить через $\sin^2 x$ и $\cos x$; отсюда

$$I = \int R_2(\sin^2 x, \cos x) \sin x\, dx = -\int R_2(1-\cos^2 x, \cos x)\, d\cos x,$$

и, если обозначить $\cos x = t$, получится интеграл от рациональной функции.

2. Аналогично, если подынтегральная функция в (33) нечетна относительно $\cos x$, то применяется подстановка $\sin x = t$.

Например,

$$\int \frac{\sin^2 x dx}{\cos^3 x - 2\sin x \cos x} = \int \frac{\sin^2 x \cos x dx}{\cos^2 x(\cos^2 x - 2\sin x)} = \left| \begin{array}{c} \sin x = t \\ \cos x\, dx = dt \end{array} \right| = \\ = \int \frac{t^2\, dt}{(1-t^2)(1-t^2-2t)};$$

полученный интеграл легко взять, разлагая подынтегральную функцию на простейшие дроби или воспользовавшись равенством

$$t^2 = \frac{1}{4}[(1-t^2)-(1-t^2-2t)]^2.$$

3. Если подынтегральная функция в (33) четна относительно $\sin x$ и $\cos x$ вместе взятых, т. е. если $R(-\sin x, -\cos x) \equiv \equiv R(\sin x, \cos x)$, то применима подстановка $\operatorname{tg} x = t$. Это нетрудно проверить в общем виде, но мы не станем этого делать, так как в каждом конкретном примере разумность подстановки видна из вычислений.

Например,

$$\int \frac{dx}{\sin^2 x \cos^4 x} = \int \frac{dx}{\operatorname{tg}^2 x \cos^6 x} = \left| \begin{array}{ll} \operatorname{tg} x = t, & \cos^2 x = \dfrac{1}{1+t^2} \\ \dfrac{dx}{\cos^2 x} = dt, & \end{array} \right| = \int \frac{(1+t^2)^2}{t^2}\, dt = \\ = \int \left(t^2 + 2 + \frac{1}{t^2}\right) dt = \frac{t^3}{3} + 2t - \frac{1}{t} + C = \operatorname{tg}^3 x + 2\operatorname{tg} x - \operatorname{ctg} x + C.$$

Этот же интеграл можно взять, представив в виде $\int \frac{(\sin^2 x + \cos^2 x)^2}{\sin^2 x \cos^4 x} dx$ и раскрыв в числителе скобки (проверьте!).

Остановимся отдельно на интегралах вида

$$\int \sin^m x \cos^n x\, dx, \tag{35}$$

где m и n — целые числа любого знака. Если m нечетное, то этот интеграл относится к разобранному случаю 1, т. е. применима подстановка $\cos x = t$. Если n нечетное, то он относится к случаю 2, а если и m и n четные, то к случаю 3. В некоторых случаях вычисления можно упростить. Например, если $m \geqslant 0$ и $n \geqslant 0$ оба четные, то применяются формулы

$$\sin^2 x = \frac{1-\cos 2x}{2}, \quad \sin x \cos x = \frac{1}{2}\sin 2x, \quad \cos^2 x = \frac{1+\cos 2x}{2}.$$

Так,

$$\int \sin^2 x \cos^6 x\, dx = \int (\sin x \cos x)^2 \cos^4 x\, dx =$$
$$= \frac{1}{16}\int \sin^2 2x\,(1+\cos 2x)^2\, dx =$$
$$= \frac{1}{16}\int \sin^2 2x\, dx + \frac{1}{8}\int \sin^2 2x \cos 2x\, dx + \frac{1}{16}\int \sin^2 2x \cos^2 2x\, dx =$$
$$= \frac{1}{32}\int (1-\cos 4x)\, dx + \frac{1}{16}\int \sin^2 2x\, d\sin 2x + \frac{1}{64}\int \sin^2 4x\, dx =$$
$$= \frac{1}{32}\left(x - \frac{\sin 4x}{4}\right) + \frac{1}{16}\frac{\sin^3 2x}{3} + \frac{1}{128}\int (1-\cos 8x)\, dx =$$
$$= \frac{5}{128}x - \frac{1}{128}\sin 4x + \frac{1}{48}\sin^3 2x - \frac{1}{1024}\sin 8x + C.$$

К такому же результату приводит выражение тригонометрических функций через показательные по формулам Эйлера (VIII.11).

К интегралам (35) применяется также интегрирование по частям с целью уменьшения положительных показателей степени и увеличения отрицательных Например,

$$\int \frac{\cos^2 x}{\sin^3 x}\, dx = \left| \begin{matrix} \cos x = u, & du = -\sin x\, dx \\ \dfrac{\cos x}{\sin^3 x} = dv, & v = -\dfrac{1}{2\sin^2 x} \end{matrix} \right| =$$
$$= -\frac{\cos x}{2\sin^2 x} - \int \frac{dx}{2\sin x} = \left| \operatorname{tg}\frac{x}{2} = t \right| =$$
$$= -\frac{\cos x}{2\sin^2 x} - \frac{1}{2}\int \frac{2dt\,(1+t^2)^{-1}}{2t\,(1+t^2)^{-1}} = -\frac{\cos x}{2\sin^2 x} - \frac{1}{2}\ln|t| + C =$$
$$= -\frac{\cos x}{2\sin^2 x} - \frac{1}{2}\ln\left|\operatorname{tg}\frac{x}{2}\right| + C$$

(при преобразовании второго интеграла мы применили формулу (34)).

11. Общие замечания. Мы видим, что интегрирование по сравнению с дифференцированием — дело гораздо более кропотливое. Поэтому хотя и нужно твердо владеть основными методами интегрирования, но нецелесообразно каждый раз, когда встречаются интегралы, заново проводить громоздкие вычисления. Имеется ряд справочников, в которых в определенном порядке собраны наиболее распространенные интегралы. Отметим, в частности, книги [6], [13], [45].

Многие важные интегралы не являются элементарными функциями, т. е. не выражаются через простейшие «школьные» функции. Например, интеграл

$$\int \sqrt[3]{x^2+1}\,dx = \int (x^2+1)^{1/3}\,dx$$

принадлежит к типу, разобранному в п. 9; но так как $n=2$, $p=\frac{1}{3}$, $m=0$, то ни один из случаев интегрируемости не имеет места. Подобным образом не являются элементарными функциями интегралы

$$\left.\begin{aligned}&\int \sin x\cdot x^\alpha\,dx,\\ &\int e^{\pm x}\cdot x^\alpha\,dx,\\ &\int \cos x\cdot x^\alpha\,dx\end{aligned}\right\}\quad (\alpha\neq 0,\ 1,\ 2,\ \ldots),$$

а потому и все интегралы, которые приводятся к этим при помощи элементарных подстановок, например,

$$\left.\begin{aligned}&\int e^{-x^2}\,dx = \left|\begin{aligned}x^2&=u\\ dx&=\frac{du}{2\sqrt{u}}\end{aligned}\right| = \frac{1}{2}\int e^{-u}u^{-1/2}\,du,\\ &\int \frac{dx}{\ln x} = |x=e^u| = \int e^u u^{-1}\,du,\\ &\int \sin x^2\,dx = |x^2=u| = \frac{1}{2}\int \sin u\cdot u^{-1/2}\,du,\end{aligned}\right\}\qquad (36)$$

$$\int \sqrt{\sin x}\,dx = \left|\begin{aligned}\sin x&=u,\\ dx&=\frac{du}{\sqrt{1-u^2}}\end{aligned}\right| = \int u^{1/2}(1-u^2)^{-1/2}\,du$$

$$\left(n=2,\ p=-\frac{1}{2},\ \ m=\frac{1}{2}\right)\ \text{и т. д.}$$

Имеются целые широкие классы неэлементарных интегралов. Например, за редкими исключениями, не берутся интегралы

$$\int R\left(x, \sqrt{P_n(x)}\right)dx, \qquad (37)$$

где R, как и раньше,— знак рациональной функции, а $P_n(x)$ — многочлен степени $n\geqslant 3$.

Когда-то тот факт, что интеграл, появляющийся при решении той или иной задачи, не берется в элементарных функциях, расценивали чуть ли не как катастрофу. Сейчас в этом нет ничего страшного. Прежде всего был введен, детально изучен и затабулирован целый ряд неэлементарных функций, через которые можно выразить наиболее важные для приложений интегралы, не выражающиеся элементарно. Например, в п. XIV.12 мы увидим, что через такие

неэлементарные (называемые *специальными*) функции легко выражаются все интегралы (36). Интеграл (37) при $n=3$ и 4 называется *эллиптическим интегралом* и выражается через так называемые эллиптические функции, также хорошо изученные. Эти выражения можно найти в указанных справочниках, к которым надо добавить книгу [53].

Кроме того, в настоящее время техника работы с интегралами так усовершенствовалась, что исследовать и применять функцию, заданную в виде интеграла, немногим сложнее, чем заданную непосредственно, без знаков интеграла. Поэтому теперь часто, даже если интеграл берется в элементарных функциях, но имеет слишком громоздкое выражение, предпочитают интеграл не брать и применять его «целиком», не взятым.

ОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ

При решении многих важных задач приходится суммировать бесконечно большое число бесконечно малых слагаемых. Это приводит к одному из центральных понятий математики, именно, к понятию определенного интеграла, ради которого излагались все методы интегрирования в гл. XIII.

§ 1. Определение и основные свойства

1. Примеры, приводящие к понятию определенного интеграла. Рассмотрим задачу, обратную той, которая разбиралась в начале п. IV.1 и привела нас к понятию производной. А именно, будем считать, что нам известен закон изменения мгновенной скорости $v=v(t)$ при движении точки и нас интересует путь, пройденный за некоторый промежуток времени от $t=\alpha$ до $t=\beta$.

Поскольку движение не предполагается равномерным, мы не можем вычислять путь как произведение скорости на истекшее время. Поэтому для подсчета пути поступим следующим образом. Разобьем весь промежуток времени на большое число малых, не обязательно равных друг другу интервалов времени

$$t_0\,(=\alpha)\leqslant t\leqslant t_1,\quad t_1\leqslant t\leqslant t_2\ ,\ \ldots,\ t_{n-1}\leqslant t\leqslant t_n\,(=\beta),$$

где $t_1, \ldots, t_{n-1}$ — некоторые промежуточные произвольно выбранные моменты времени. Если эти интервалы достаточно малы, то без большой ошибки на каждом из них движение можно считать равномерным, что даст приближенное выражение для пути

$$s\approx v_1\,\Delta t_1+v_2\,\Delta t_2+\ldots+v_n\,\Delta t_n; \tag{1}$$

здесь $\Delta t_k=t_k-t_{k-1}$, а v_k — какое-либо из значений мгновенной скорости v на k-м интервале времени, т е. $v_k=v(\tau_k), t_{k-1}\leqslant\tau_k\leqslant t_k$. (Обратим внимание читателя на некоторое отличие этих обозначений от обозначений § V.2, где было $t_k-t_{k-1}=\Delta t_{k-1}$, $v_k=v(t_k)$.) Поэтому формулу (1) можно иначе записать так:

$$s\approx\sum_{k=1}^{n}v(\tau_k)\,\Delta t_k\qquad(\alpha=t_0<t_1<\ldots<t_n=\beta;\quad t_{k-1}\leqslant\tau_k\leqslant t_k).$$

Эта формула тем точнее, чем мельче разбиение основного промежутка времени; чтобы получить точную формулу, *надо перейти*

к пределу, приняв, что это разбиение бесконечно измельчается:

$$s = \lim \sum_{k=1}^{n} v(\tau_k)\,\Delta t_k. \qquad (2)$$

Подобным образом во втором примере п. IV.1, в задаче о наполнении сосуда, при известной, но переменной скорости наполнения $w = w(t)$ общий объем V, наполненный за промежуток времени от α до β, равен

$$V = \lim \sum_{k=1}^{n} w(\tau_k)\,\Delta t_k \qquad (3)$$

при том же смысле обозначений. Как и для формулы (2), это основано на том, что на протяжении малого промежутка времени скорость наполнения при подсчете наполненного объема можно считать почти постоянной; точнее говоря, на протяжении бесконечно малого промежутка времени эту скорость можно считать постоянной.

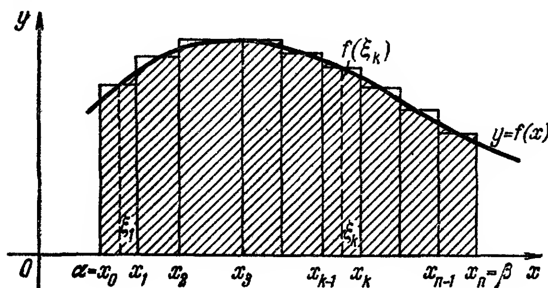


Рис. 244.

В третьем примере п. IV.1, если считать заданной линейную плотность ρ нити в каждой точке s, т. е. $\rho = \rho(s)$, аналогично получим общую массу нити

$$M = \lim \sum_{k=1}^{n} \rho(\sigma_k)\,\Delta s_k \quad (\alpha = s_0 < s_1 < \ldots < s_n = \beta,\ s_{k-1} \leqslant \sigma_k \leqslant s_k), \qquad (4)$$

причем предел берется в процессе, когда воображаемое разбиение нити бесконечно измельчается; α и β — значения s, отвечающие концам нити.

Рассмотрим, наконец, важный геометрический пример. Пусть требуется вычислить площадь фигуры, заштрихованной на рис. 244 и называемой *криволинейной трапецией*, причем для простоты будем считать, что $f(x) > 0$. Если разбить весь интервал $\alpha \leqslant x \leqslant \beta$ изменения x на малые промежутки при помощи точек деления $x_0 = \alpha < x_1 < x_2 < \ldots < x_{n-1} < x_n = \beta$ и принять высоту на каждом

из этих малых промежутков постоянной, то получим приближенное выражение для площади криволинейной трапеции

$$S \approx \sum_{k=1}^{n} f(\xi_k)\,\Delta x_k \qquad (\Delta x_k = x_k - x_{k-1},\ x_{k-1} \leqslant \xi_k \leqslant x_k).$$

Геометрический смысл правой части — это площадь ступенчатой фигуры, изображенной на рис. 244 и полученной в результате замены каждого из n столбиков, на которые разбита криволинейная трапеция, прямоугольником с тем же основанием и с высотой, равной одной из высот столбика. В пределе, при бесконечном измельчении разбиения, получаем

$$S = \lim \sum_{k=1}^{n} f(\xi_k)\,\Delta x_k. \tag{5}$$

2. Основное определение. Выражения (2) — (5), получающиеся при решении различных задач, имеют одинаковую структуру. Аналогичные выражения получаются и во многих других задачах, что дает основание для следующего общего определения.

Пусть некоторая функция $f(x)$ задана при $\alpha \leqslant x \leqslant \beta$. Разобьем произвольно этот интервал на маленькие промежутки при помощи точек деления $x_0 = \alpha < x_1 < x_2 < \ldots < x_n = \beta$ и образуем *интегральную сумму*

$$\sum_{k=1}^{n} f(\xi_k)\,\Delta x_k = f(\xi_1)\,\Delta x_1 + f(\xi_2)\,\Delta x_2 + \ldots + f(\xi_n)\,\Delta x_n, \tag{6}$$

где $\Delta x_k = x_k - x_{k-1}$, а каждая из точек ξ_k произвольно выбрана между x_{k-1} и x_k, т. е. на k-м промежутке разбиения. *Пусть теперь разбиение бесконечно измельчается; тогда предел, к которому стремится интегральная сумма в этом процессе, называется определенным интегралом от функции $f(x)$ по интервалу интегрирования* $\alpha \leqslant x \leqslant \beta$ и обозначается

$$\int_{\alpha}^{\beta} f(x)\,dx = \lim \sum_{k=1}^{n} f(\xi_k)\,\Delta x_k. \tag{7}$$

В примерах п. 1 получаем соответственно

$$s = \int_{\alpha}^{\beta} v(t)\,dt, \quad V = \int_{\alpha}^{\beta} w(t)\,dt, \quad M = \int_{\alpha}^{\beta} \rho(s)\,ds, \quad S = \int_{\alpha}^{\beta} f(x)\,dx. \tag{8}$$

Из последнего равенства ясен геометрический смысл определенного интеграла в случае, если подынтегральная функция положительна: *он равен площади криволинейной трапеции, ограниченной графиком подынтегральной функции, осью абсцисс и прямыми, параллельными оси ординат, построенными на концах интервала интегрирования.*

Эти концы называются *нижним* и *верхним пределами интегрирования*.

Если подынтегральная функция отрицательна или меняет знак, то в интегральную сумму (6) некоторые члены войдут со знаком минус. В пределе получится, что интеграл равен *алгебраической сумме площадей* участков криволинейной трапеции, причем (рис. 245) площади участков, лежащих выше (соответственно ниже) оси x, берутся со знаком $+$ (соответственно $-$).

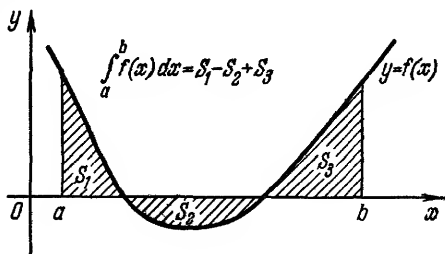


Рис. 245.

Из сравнения формул (8) следует также, что для получения пройденного пути при заданной на графике (рис. 246) зависимости скорости от времени надо просто взять площадь соответствующей криволинейной трапеции. И здесь, если $v < 0$, график проходит ниже оси t, а приращение координаты движущейся точки отрицательно, т. е. указанную площадь надо брать со знаком $-$. Это *правило знаков при подсчете площадей* действует и в большом числе других примеров.

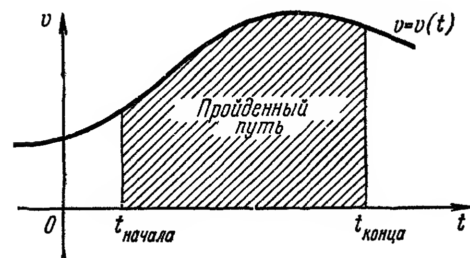


Рис. 246.

Остановимся на предельном переходе в формуле (7). Иногда говорят, что предел берется при $n \to \infty$; это не совсем точно, так как участки Δx_k не предполагаются равными и если потребовать только, чтобы $n \to \infty$, то может получиться, что одна часть интервала $\alpha \leqslant x \leqslant \beta$ измельчается, а другая — нет. Лучше говорить, что *предел берется при безграничном измельчении интервала интегрирования*. Степень этого измельчения можно охарактеризовать наибольшей из длин Δx_k участков данного разбиения, так как если наибольшая из этих длин мала, то и все длины малы. Поэтому можно сказать, что предельный переход в формуле (7) осуществляется в процессе, когда $\max\limits_k \Delta x_k \to 0$.

Приведем пример подсчета интеграла на основе его определения (7). Пусть надо вычислить интеграл $\int\limits_0^1 x^2\,dx$. Разобьем интервал интегрирования на пять равных частей длины 0,2 и выберем для опреде-

ленности на каждой из этих частей точку в левом конце. Тогда $\xi_1 = 0{,}0$; $\xi_2 = 0{,}2$; $\xi_3 = 0{,}4$; $\xi_4 = 0{,}6$; $\xi_5 = 0{,}8$ и

$$\int_0^1 x^2\,dx \approx \sum_{k=1}^{5} \xi_k^2\,\Delta x_k = (0{,}0^2 + 0{,}2^2 + 0{,}4^2 + 0{,}6^2 + 0{,}8^2)\cdot 0{,}2 = 0{,}24.$$

При аналогичном разбиении на 10 частей получилось бы значение 0,29, а при разбиении на 100 частей — значение 0,33, которое совсем близко к точному значению, равному $\frac{1}{3}$, как мы увидим в п. 3. (Разберите на чертеже, почему приближенные значения в данном примере получились меньше точного!)

Таким образом, хотя интегральная сумма (6) обладает высокой степенью произвола, она зависит как от выбора точек деления x_i, так и от выбора промежуточных точек ξ_k, но если разбиение взять достаточно мелким, то эта сумма практически просто равна своему пределу, т. е. интегралу (7) (который, конечно, уже не зависит ни от точек x_k, ни от ξ_k). При достаточно мелком разбиении каждое из слагаемых в сумме становится весьма малым за счет малости Δx_k, но число слагаемых столь велико, что вся сумма имеет конечное значение; грубо говоря, если число слагаемых в сумме равно n, то все Δx_k, а потому и каждое слагаемое в сумме имеют порядок $\frac{1}{n}$. Учитывая знак предела в формуле (7), можно сказать, что *определенный интеграл — это сумма бесконечно большого числа бесконечно малых слагаемых.* На практике часто можно определенный интеграл рассматривать просто как сумму весьма большого числа весьма малых однородных (т. е. одной размерности, одного характера, одного смысла) слагаемых, столь малых, что эта сумма практически равна своему пределу. Это вполне соответствует понятию о практических бесконечно больших и бесконечно малых величинах, как о величинах достаточно больших и достаточно малых, но теоретически конечных (см. п. III.1 и III.3). Отметим, что не всякая сумма бесконечно большого числа бесконечно малых слагаемых дает интеграл: это ясно уже из того, что, как было сказано выше, число слагаемых и их размер должны быть согласованы так, чтобы эта сумма принимала конечное значение.

Из такого истолкования интеграла как суммы проистекло его обозначение. Если считать слагаемые в сумме (6) практически бесконечно малыми и обозначить тогда $\Delta x_k = dx$, то всю сумму (6) можно переписать в виде $\sum\limits_{\text{от } x=\alpha}^{\text{до } x=\beta} f(x)\,dx$. В старину сумму обозначали буквой S, от латинского summa; эта буква постепенно удлинялась, что привело к современному обозначению (7).

В заключение отметим, что подынтегральная функция на интервале интегрирования может быть либо непрерывной, либо разрывной,

т. е. иметь точки разрыва. Однако в этом параграфе обязательно требуется, *чтобы интервал интегрирования был конечным и подынтегральная функция на этом интервале нигде не обращалась в бесконечность*. Тогда в более полных курсах математического анализа доказывается, без всякой ссылки на геометрический или физический смысл, что интеграл обязательно существует, т. е. имеет определенное конечное числовое значение. Если же указанные условия нарушаются, то, как будет показано в § 4, интеграл может оказаться лишенным численного значения.

3. Связь определенного интеграла с неопределенным. Начнем с простого замечания: *определенный интеграл не зависит от обозначения переменной интегрирования*, т. е.

$$\int_{\alpha}^{\beta} f(x)\,dx = \int_{\alpha}^{\beta} f(t)\,dt = \int_{\alpha}^{\beta} f(s)\,ds = \ldots \tag{9}$$

Это вытекает хотя бы из геометрического смысла интеграла, так как все выписанные интегралы равны одной и той же площади. Итак, *переменная интегрирования в определенном интеграле является немой*, подобно индексу суммирования (см. п. III. 6), и потому может быть обозначена любой буквой.

Пусть дана некоторая функция $f(x)$, которую мы будем интегрировать. Однако только нижний предел x_0 мы будем считать зафиксированным, а верхний предел x будем считать произвольным, переменным. Тогда результат интегрирования будет зависеть от x и мы его обозначим через $\Phi(x)$. Это можно записать так: $\Phi(x) = \int_{x_0}^{x} f(x)\,dx$ $(x_0 = \text{const})$ или, лучше, имея в виду равенства (9), так:

$$\Phi(x) = \int_{x_0}^{x} f(t)\,dt \qquad (x_0 = \text{const}). \tag{10}$$

Первая запись, хотя и допустимая, иногда приводит к недоразумениям, так как буква x в ней имеет два различных смысла (верхний предел и переменная интегрирования), о чем надо помнить.

Докажем, что построенная функция $\Phi(x)$ является первообразной (п. XIII.1) для подынтегральной функции $f(x)$, т. е.

$$\frac{d}{dx}\int_{x_0}^{x} f(t)\,dt = f(x):$$

производная от интеграла по верхнему пределу равна значению подынтегральной функции на верхнем пределе. Для этого сначала предположим функцию $f(x)$ непрерывной и рассмотрим рис. 247. Из геометрического смысла интеграла вытекает, что если x получит

приращение Δx, то $\Delta\Phi$ равно заштрихованной площади. Эта площадь равна произведению $\Delta x \cdot f^*$, где f^* — некоторая средняя ордината между x и $x+\Delta x$. Отсюда $\frac{\Delta\Phi}{\Delta x}=f^*=f(x^*)$, и если $\Delta x \to 0$, то $x^* \to x$ и в пределе получаем

$$\Phi'(x)=\lim_{\Delta x \to 0} \frac{\Delta\Phi}{\Delta x}=\lim_{\Delta x \to 0} f(x^*)=f(x),$$

что и требовалось доказать. Мы видим, в частности, что *непрерывная функция всегда имеет первообразную* (см. п. XIII.1); чтобы ее получить, надо взять определенный интеграл от заданной функции при зафиксированном нижнем пределе и рассмотреть его как функцию от верхнего предела.

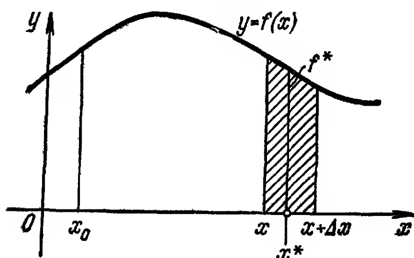


Рис. 247.

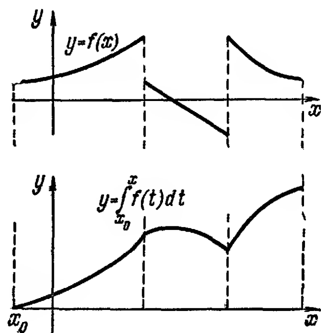


Рис. 248.

Если подынтегральная функция разрывна (но конечна, так как мы пока рассматриваем только конечные функции), то в ее точках разрыва функция (10) непрерывна, но имеет «излом» (рис. 248), так как производная от $\Phi(x)$ при переходе через такую точку должна претерпеть скачок. Допуская такие изломы, мы расширяем понятие первообразной функции, так как в самой точке излома единой производной нет. При этом естественном расширении получается, что и всякая всюду конечная функция имеет первообразную, которая является непрерывной функцией.

Допустим теперь, что нам надо вычислить интеграл

$$I=\int_{\alpha}^{\beta} f(x)\,dx$$

и мы знаем *какую-то* первообразную $F(x)$ к функции $f(x)$. Так как функция $\int_{\alpha}^{x} f(t)\,dt$ — тоже первообразная к $f(x)$, то в силу п. XIII.1

$$\int_{\alpha}^{x} f(t)\,dt=F(x)+C,$$

где C — некоторая постоянная. Если здесь положить $x=\alpha$, то из геометрического смысла интеграла вытекает, что левая часть обратится в нуль, т. е.

$$0=F(\alpha)+C; \quad C=-F(\alpha), \ \int_{\alpha}^{x} f(t)\,dt=F(x)-F(\alpha).$$

Если в последней формуле положить $x=\beta$, то на основе (9) получим

$$\int_{\alpha}^{\beta} f(x)\,dx=F(\beta)-F(\alpha), \tag{11}$$

т. е. *определенный интеграл равен приращению первообразной к подынтегральной функции, когда независимая переменная изменяется от нижнего до верхнего предела.* Правую часть равенства (11) записывают еще в виде $F(x)\Big|_{\alpha}^{\beta}$, где $\Big|_{\alpha}^{\beta}$ — знак *двойной подстановки*, который означает, что в рассматриваемую функцию надо подставить вместо аргумента верхний предел, затем нижний и из первого результата вычесть второй.

Формулу (11) записывают еще так:

$$\int_{\alpha}^{\beta} f(x)\,dx=\left(\int f(x)\,dx\right)\Big|_{\alpha}^{\beta}, \tag{12}$$

так как (см. п. XIII.1)

$$\left(\int f(x)\,dx\right)\Big|_{\alpha}^{\beta}=(F(x)+C)\Big|_{\alpha}^{\beta}=$$
$$=[F(\beta)+C]-[F(\alpha)+C]=F(\beta)-F(\alpha)=\int_{\alpha}^{\beta} f(x)\,dx.$$

Итак, *определенный интеграл равен приращению неопределенного.* Этот результат, один из важнейших в математике, называется *теоремой Ньютона — Лейбница.*

Например,

$$\int_{0}^{1} x^2\,dx=\left(\int x^2\,dx\right)\Big|_{0}^{1}=\left(\frac{x^3}{3}\right)\Big|_{0}^{1}=\frac{1^3}{3}-\frac{0^3}{3}=\frac{1}{3}.$$

Отметим, что здесь мы при вычислении неопределенного интеграла не писали произвольной постоянной C, так как выше мы видели, что члены $+C$ и $-C$ все равно взаимно уничтожаются. Хорошо видно, что *определенный интеграл при заданных пределах интегрирования является постоянным числом, тогда как неопределенный интеграл является функцией.*

До сих пор мы считали, что $\alpha<\beta$. *Если* $\alpha\geqslant\beta$, *то формула* (11) *принимается за определение интеграла, стоящего в левой части.*

Так как $f(x)=F'(x)$, то формулу (11) можно переписать так:

$$\int_{\alpha}^{\beta} F'(x)\,dx=F(\beta)-F(\alpha), \tag{13}$$

т. е. *определенный интеграл от производной равен приращению первообразной.*

4. Основные свойства определенного интеграла. 1. *При перестановке пределов интегрирования интеграл умножается на* -1. Действительно, в силу формулы (11)

$$\int_{\beta}^{\alpha} f(x)\,dx=F(\alpha)-F(\beta)=-[F(\beta)-F(\alpha)]=-\int_{\alpha}^{\beta} f(x)\,dx.$$

Это простое свойство, которое можно записать также в виде $F(x)\Big|_{\alpha}^{\beta}=-F(x)\Big|_{\beta}^{\alpha}$, дает возможность, изменив знак у неопределенного интеграла, подставлять пределы в обратном порядке, т. е. вычислять так:

$$\int_{3}^{5} \frac{1}{x^2}\,dx=-\frac{1}{x}\Big|_{3}^{5}=\frac{1}{3}-\frac{1}{5}=\frac{2}{15}.$$

Из свойства 1, в частности, вытекает правило дифференцирования интеграла по нижнему пределу:

$$\frac{d}{dx}\left(\int_{x}^{x_0} f(t)\,dt\right)=-\frac{d}{dx}\left(\int_{x_0}^{x} f(t)\,dt\right)=-f(x).$$

Следующее свойство уже применялось в п. 3.

2. *Если пределы интегрирования совпадают, то интеграл равен нулю,* т. е.

$$\int_{\alpha}^{\alpha} f(x)\,dx=0.$$

3. *«Теорема о разбиении интервала интегрирования»*:

$$\int_{\alpha}^{\beta} f(x)\,dx+\int_{\beta}^{\gamma} f(x)\,dx=\int_{\alpha}^{\gamma} f(x)\,dx.$$

Действительно, левая часть равна

$$[F(\beta)-F(\alpha)]+[F(\gamma)-F(\beta)]=F(\gamma)-F(\alpha)=\int_{\alpha}^{\gamma} f(x)\,dx.$$

4. *Интеграл от суммы равен сумме интегралов*:

$$\int_{\alpha}^{\beta} [f(x)\pm\varphi(x)]\,dx=\int_{\alpha}^{\beta} f(x)\,dx\pm\int_{\alpha}^{\beta} \varphi(x)\,dx.$$

Для доказательства надо исходить из аналогичного свойства (XIII.9) неопределенных интегралов и приравнять приращения левой и правой частей, когда x меняется от α до β. Аналогично доказывается свойство

5. *Постоянный множитель можно выносить за знак интеграла:*

$$\int_\alpha^\beta Mf(x)\,dx = M\int_\alpha^\beta f(x)\,dx \qquad (M=\text{const}).$$

Свойства 4 и 5 совместно читаются так: *определенный интеграл линеен относительно подынтегральной функции.*

Здесь термин «линеен» понимается в том самом смысле, как в п. XI.6. Именно, формула

$$\int_\alpha^\beta f(x)\,dx = I \tag{14}$$

при зафиксированных α, β ставит в соответствие каждой конечной функции, определенной при $\alpha \leqslant x \leqslant \beta$, число I. Другими словами, формула (14) определяет отображение бесконечномерного линейного пространства всех таких функций в одномерное линейное пространство всех чисел, а свойства 4 и 5 означают, что это отображение линейное. (Проверьте, например, что при $\alpha=1$, $\beta=2$ функции $y=x^2$ отвечает число $I=\frac{7}{3}$, функции $y=\frac{1}{x^3}$ — число $\frac{3}{8}$, а функции $y=5x^2-\frac{3}{x^3}$ — число $5\cdot\frac{7}{3}-3\cdot\frac{3}{8}=10{,}54$.) Закон, по которому функциям ставятся в соответствие числа, называется *функционáлом*, значит, формула (14) определяет *линейный функционал*, определенный на указанном пространстве функций

6. *Формула интегрирования по частям*

$$\int_\alpha^\beta uv'\,dx = (uv)\Big|_{x=\alpha}^{\beta} - \int_\alpha^\beta u'v\,dx$$

также получается из соответствующей формулы для неопределенных интегралов, а именно, формулы (XIII.13).

Например,

$$\int_0^\pi x\sin x\,dx = \begin{vmatrix} u=x, & dv=\sin x\,dx \\ du=dx, & v=-\cos x \end{vmatrix} =$$

$$=(-x\cos x)\Big|_0^\pi + \int_0^\pi \cos x\,dx = (-x\cos x)\Big|_0^\pi + (\sin x)\Big|_0^\pi = \pi.$$

7. *Формула замены переменной для определенных интегралов* получается, если приравнять приращения обеих частей формулы (XIII.15), когда t меняется от α до β. Учитывая, что при этом x,

равный $\varphi(t)$, изменится от $\varphi(\alpha)$ до $\varphi(\beta)$, получим

$$\int_{\alpha}^{\beta} f[\varphi(t)]\,\varphi'(t)\,dt = \left[\int f(x)\,dx\right]\Big|_{x=\varphi(\beta)} - \left[\int f(x)\,dx\right]\Big|_{x=\varphi(\alpha)}$$

или, учитывая формулу (12),

$$\int_{\alpha}^{\beta} f[\varphi(t)]\,\varphi'(t)\,dt = \int_{\varphi(\alpha)}^{\varphi(\beta)} f(x)\,dx.$$

Таким образом, здесь надо дополнительно произвести замену пределов интегрирования, выяснив интервал, который должна пройти новая переменная, чтобы связанная с ней старая переменная интегрирования прошла первоначально заданный для нее интервал.

Если, например, при вычислении интеграла

$$\int_0^R \sqrt{R^2-x^2}\,dx$$

мы хотим совершить подстановку $x=R\sin t$, то надо учесть, что если x менялось от 0 до R, то t уже будет меняться от 0 до $\frac{\pi}{2}$. Поэтому

$$\int_0^R \sqrt{R^2-x^2}\,dx = \int_0^{\frac{\pi}{2}} \sqrt{R^2-R^2\sin^2 t}\;R\cos t\,dt =$$

$$= R^2\int_0^{\frac{\pi}{2}} \cos^2 t\,dt = \frac{R^2}{2}\int_0^{\frac{\pi}{2}} (1+\cos 2t)\,dt = \frac{R^2}{2}\left(t+\frac{\sin 2t}{2}\right)\Big|_0^{\frac{\pi}{2}} = \frac{\pi R^2}{4}.$$

Как видим, в отличие от замены переменной в неопределенном интеграле, здесь не требуется выполнять обратную подстановку, т. е. перейти в ответе к старой переменной. Мы предоставляем читателю выяснить, площадь какой фигуры только что найдена, а также получить тот же результат с помощью замены $x=R\cos t$.

Мы вывели свойства 3—5 определенного интеграла с помощью формулы (11). *Однако их можно было бы вывести и на основе определения* (7) *интеграла как предела интегральной суммы*. Например, переходя в формуле

$$\sum_{k=1}^{n} [f_1(\xi_k) \pm f_2(\xi_k)]\,\Delta x_k = \sum_{k=1}^{n} f_1(\xi_k)\,\Delta x_k \pm \sum_{k=1}^{n} f_2(\xi_k)\,\Delta x_k$$

к пределу при бесконечном измельчении разбиения интервала $\alpha \leqslant x \leqslant \beta$, получаем свойство 4 и т. д. Из того же определения вытекает свойство

8. Если рассматриваемые переменные размерны, то

$$\left[\int_{\alpha}^{\beta} f(x)\,dx\right] = [f]\cdot[x],$$

так как суммирование и переход к пределу не меняют размерности.

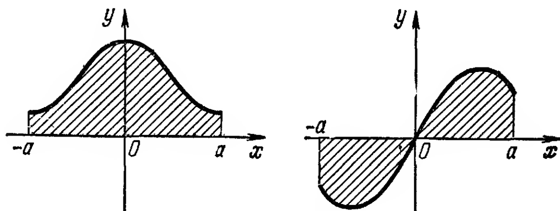


Рис. 249.

9. *Интегрирование в симметричных пределах часто можно упростить* по формулам (рис. 249):

$$\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx, \text{ если } f(x) \text{ — четная функция,}$$

$$\int_{-a}^{a} f(x)\,dx = 0, \text{ если } f(x) \text{ — нечетная функция.}$$

10. *Интеграл от периодической функции по периоду не зависит от положения интервала интегрирования*: другими словами, если $f(x+A)\equiv f(x)$, то интеграл

$$I = \int_{x}^{x+A} f(s)\,ds$$

не зависит от x. В самом деле, по правилу дифференцирования сложной функции и на основе формул для производной интеграла по верхнему и нижнему пределам получаем

$$\frac{dI}{dx} = f(x+A)\,\frac{d(x+A)}{dx} - f(x)\,\frac{dx}{dx} = f(x+A) - f(x) = 0.$$

(Докажите это же свойство, исходя из геометрического смысла определенного интеграла.)

Приведем в заключение несколько примеров ошибочного вычисления интегралов.

1. $$\int_{-1}^{1} \sqrt{1-x^2}\,dx = \left|\begin{matrix} x=\cos t \\ \pi\leqslant t\leqslant 2\pi \end{matrix}\right| = \int_{\pi}^{2\pi} \sqrt{1-\cos^2 t}\,(-\sin t)\,dt =$$

$$= -\int_{\pi}^{2\pi} \sin^2 t\,dt = -\int_{\pi}^{2\pi} \frac{1-\cos 2t}{2}\,dt = -\frac{1}{2}\left(t - \frac{\sin 2t}{2}\right)\Big|_{\pi}^{2\pi} = -\frac{\pi}{2};$$

этот результат заведомо ошибочный, так как *интеграл от положительной функции в положительном направлении*, т. е. от меньшего к большему, *должен быть положительным*. Ошибка состояла в том, что мы заменили $\sqrt{\sin^2 t}$ на $\sin t$, тогда как надо было на $|\sin t|$ (см. конец п. 1.5); а так как при $\pi < t < 2\pi$ будет $\sin t < 0$, то для таких t будет $\sqrt{\sin^2 t} = -\sin t$, что привело бы к правильному результату.

В аналогичных случаях иногда получаются интегралы вида

$$\int_a^b |f(x)|\,dx.$$

Такие интегралы можно вычислять так. Находим интервалы знакопостоянства функции $f(x)$ (п. III.15); пусть, например, $f(x) > 0$ при $a < x < c$, $f(x) < 0$ при $c < x < d$ и $f(x) > 0$ при $d < x < b$. Тогда

$$\int_a^b |f(x)|\,dx = \int_a^c |f(x)|\,dx + \int_c^d |f(x)|\,dx + \int_d^b |f(x)|\,dx =$$
$$= \int_a^c f(x)\,dx - \int_c^d f(x)\,dx + \int_d^b f(x)\,dx$$

и т. д.

2. $$\int_{-1}^{2} x^2 dx = \left| \begin{matrix} x^2 = t \\ x = \sqrt{t} \end{matrix} \right| = \int_1^4 t\,\frac{1}{2\sqrt{t}}\,dt =$$
$$= \frac{1}{2}\int_1^4 \sqrt{t}\,dt = \frac{1}{2}\,\frac{t^{3/2}}{\frac{3}{2}}\Bigg|_1^4 = \frac{8}{3} - \frac{1}{3} = \frac{7}{3};$$

это противоречит правильному значению 3, которое получится, если не пользоваться подстановкой. Ошибка состояла в том, что формула перехода $x = \sqrt{t}$ при $x < 0$ недействительна. Если почему-либо надо провести указанную подстановку $x^2 = t$, то надо разбить $\int_{-1}^{2} x^2\,dx = \int_{-1}^{0} + \int_0^2$ и в первом интеграле положить $x = -\sqrt{t}$ $(1 \geqslant t \geqslant 0)$, а во втором $x = \sqrt{t}$ $(0 \leqslant t \leqslant 4)$ (проделайте это!).

3. $$\int_{-1}^{2} x^{-2}\,dx = \frac{x^{-1}}{-1}\Bigg|_{-1}^{2} = \frac{2^{-1} - (-1)^{-1}}{-1} = -\frac{3}{2}.$$

Этот результат, как и в примере 1, заведомо ошибочный. Ошибка состояла в том, что подынтегральная функция обращается на интервале интегрирования в бесконечность при $x = 0$. Как быть с такими интегралами, будет сказано в п. 16

5. Интегрирование неравенств. Из определения интеграла и из его геометрического смысла (п. 2) вытекает, что

$$\text{если } f(x) \geqslant 0 \text{ и } \alpha < \beta, \text{ то } \int_\alpha^\beta f(x)\,dx \geqslant 0. \qquad (15)$$

Последнее неравенство может обратиться в равенство, только если $f(x) \equiv 0$ $(\alpha \leqslant x \leqslant \beta)$. Впрочем, подынтегральная функция может быть отличной от нуля в отдельных, дискретных точках, так как такие значения не сказываются на значении интеграла.

Если дано, что

$$\varphi(x) \leqslant \psi(x) \quad \text{при} \quad \alpha \leqslant x \leqslant \beta, \tag{16}$$

то, обозначив $\psi(x) - \varphi(x) = f(x)$ и применяя утверждение (15), получим

$$\int_\alpha^\beta [\psi(x) - \varphi(x)]\,dx \geqslant 0, \quad \int_\alpha^\beta \psi(x)\,dx - \int_\alpha^\beta \varphi(x)\,dx \geqslant 0,$$

т. е.

$$\int_\alpha^\beta \varphi(x)\,dx \leqslant \int_\alpha^\beta \psi(x)\,dx. \tag{17}$$

Итак, из неравенства (16) мы получили (17), т. е. *неравенства можно интегрировать в положительном направлении*. (Продумайте, что произойдет при интегрировании неравенства в отрицательном направлении.)

Как и выше, в условиях (16) неравенство (17) может обратиться в равенство, только если $\varphi(x) \equiv \psi(x)$ при $\alpha \leqslant x \leqslant \beta$.

В качестве следствия получим *самую грубую оценку интеграла*. Пусть

$$f_{\min} \leqslant f(x) \leqslant f_{\max} \qquad (\alpha \leqslant x \leqslant \beta),$$

где $f_{\min}$ и $f_{\max}$ — две константы. Интегрируя эти неравенства, получим

$$f_{\min}(\beta - \alpha) \leqslant \int_\alpha^\beta f(x)\,dx \leqslant f_{\max}(\beta - \alpha). \tag{18}$$

В связи с этой оценкой находится важное понятие *среднего* (говорят также «среднего интегрального» или «среднего арифметического») *значения функции*. Если функция $f(x)$ рассматривается на интервале $\alpha \leqslant x \leqslant \beta$, то ее средним значением на этом интервале называется такая константа $\bar{f}$, интеграл от которой по интервалу $\alpha \leqslant x \leqslant \beta$ равен интегралу от функции $f(x)$ по этому интервалу. Таким образом,

$$\int_\alpha^\beta \bar{f}\,dx = \int_\alpha^\beta f(x)\,dx, \quad \int_\alpha^\beta f(x)\,dx = \bar{f}\cdot(\beta - \alpha).$$

Из последней *формулы среднего значения* получаем выражение для среднего значения

$$\bar{f}=\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} f(x)\,dx. \tag{19}$$

Из неравенства (18) вытекает, как и следует ожидать, что

$$f_{\min}\leqslant\bar{f}\leqslant f_{\max}.$$

Геометрический смысл среднего значения функции показан на рис. 250: $\bar{f}$ должно быть таким, чтобы площадь прямоугольника $AB'C'D$ равнялась площади криволинейной трапеции $ABCD$. Ясно, что если функция $f(x)$ непрерывная, то она принимает где-то на интервале $\alpha<x<\beta$ значение $\bar{f}$ (на рис. 250 при $x=\gamma$); разрывная функция может не принимать своего среднего значения.

Разумность приведенного здесь определения среднего значения функции хорошо видна, например, при рассмотрении зависимости мгновенной скорости от времени в процессе неравномерного движения точки. Так как интеграл от скорости по времени равен пути (см. первую формулу 8)), то получается, что среднее значение скорости за некоторый промежуток времени — это такая постоянная скорость, при которой точка за тот же промежуток времени прошла бы тот же путь, что и при рассматриваемом неравномерном движении. Другими словами, в силу (19) среднее значение скорости за конечный промежуток времени при неравномерном движении — это отношение пройденного пути к истекшему времени; таким образом, это понятие совпадает с хорошо известным понятием средней скорости. Применяющиеся в физике понятия средней плотности, средней мощности и т. д. также находятся в соответствии с указанным общим понятием среднего значения функции.

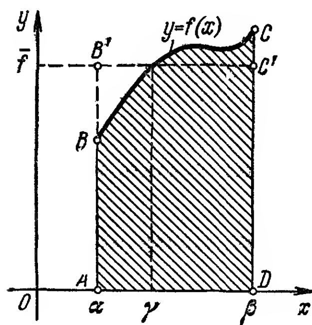


Рис. 250.

Если функция $f(x)$ задана на бесконечном интервале, например, при $\alpha\leqslant x<\infty$, то ее средним значением называется

$$\bar{f}=\lim_{\beta\to\infty}\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} f(x)\,dx,$$

т. е. предел среднего значения по конечному интервалу. Легко проверить, что этот предел, если он существует, не зависит от выбора значения α.

Например, в цепи переменного тока сила тока и напряжение обычно выражаются формулами $j=j_0\cos(\omega t+\alpha)$, $u=u_0\cos(\omega t+\alpha+\varphi)$, где φ — постоянный сдвиг фазы напряжения по сравнению с силой тока. Поэтому средняя

потребляемая мощность равна

$$h=\overline{ju}=\lim_{T\to\infty}\frac{1}{T}\int_0^T j_0\cos(\omega t+\alpha)\,u_0\cos(\omega t+\alpha+\varphi)\,dt=$$

$$=\lim_{T\to\infty}\frac{j_0u_0}{2T}\int_0^T[\cos(2\omega t+2\alpha+\varphi)+\cos\varphi]\,dt=$$

$$=\lim_{T\to\infty}\left\{\frac{j_0u_0}{4\omega}\,\frac{\sin(2\omega T+2\alpha+\varphi)-\sin(2\alpha+\varphi)}{T}+\frac{j_0u_0}{2}\cos\varphi\right\}=\frac{j_0u_0}{2}\cos\varphi.$$

Отсюда вытекает значение величины $\cos\varphi$ в электротехнике.

Укажем в заключение одно неравенство, которое иногда применяется. Если для интегральной суммы (6) записать, что абсолютное значение суммы не превосходит суммы абсолютных значений (см. конец п. 1.5):

$$\left|\sum_{k=1}^{n}f(\xi_k)\,\Delta x_k\right|\leqslant\sum_{k=1}^{n}|f(\xi_k)\,\Delta x_k|=\sum_{k=1}^{n}|f(\xi_k)|\,\Delta x_k,$$

а затем перейти к пределу, то мы получим

$$\left|\int_a^b f(x)\,dx\right|\leqslant\int_a^b|f(x)|\,dx. \qquad (20)$$

Другими словами, *абсолютное значение интеграла не превосходит интеграла от абсолютного значения функции* (подумайте, когда неравенство (20) обращается в равенство).

§ 2. Применения определенного интеграла

6. Две схемы применения. Имеются две основные схемы применения определенного интеграла к нахождению геометрических, физических и т. п. величин.

Первая схема основана на определении интеграла как предела интегральной суммы (см. формулу (7)). Изучаемая величина приближенно представляется в виде интегральной суммы, причем с измельчением разбиения это представление становится все более точным и в пределе переходит в точное. Поэтому данная величина равна пределу интегральной суммы, т. е. интегралу. Этот прием достаточно продемонстрирован на примерах п. 1, которые привели к четырем интегралам (8). Как указано в п. 2, эта схема основана на представлении об интеграле как о сумме бесконечно большого числа бесконечно малых слагаемых.

Вторая схема применения интегралов состоит в том, что составляется соотношение между дифференциалами рассматриваемых величин, так называемое *дифференциальное уравнение*. От этого

соотношения мы переходим к соотношению между самими величинами при помощи интегрирования на основе формулы (13), которую можно записать также в виде

$$\int dy = y_{\text{конечное}} - y_{\text{начальное}}.$$

Это соотношение означает просто, что сумма всех бесконечно малых приращений какой-либо величины равна полному приращению этой величины.

Рассмотрим пример. Пусть точка движется вдоль оси s, проходя отрезок от $s=a$ до $s=b$, причем на нее действует переменная сила, направленная вдоль оси и принимающая в каждой точке s значение $F(s)$. Пусть требуется вычислить общую работу $A_{\text{общ}}$, произведенную этой силой на указанном пути. Для этого заметим, что работа A, произведенная силой в процессе движения, находится в функциональной связи с пройденным путем, т. е. $A=A(s)$. Если точка проходит малый интервал от s до $s+\Delta s$, то сила не успевает существенно измениться и ее можно приближенно считать постоянной, т. е. написать по известной формуле физики

$$\Delta A \approx F(s)\,\Delta s.$$

Более точная формула имеет вид

$$\Delta A = F(s)\,\Delta s + \alpha, \text{ где } |\alpha| \ll \Delta s, \qquad (21)$$

т. е. α имеет высший порядок малости, чем Δs. То, что здесь действительно высший порядок малости, вытекает из следующего рассуждения: α получается из-за того, что F на участке Δs успевает измениться, но при бесконечно малом Δs это изменение также бесконечно мало́, а при подсчете ΔA это изменение надо еще множить на Δs.

Если теперь вспомнить определение дифференциала как главной линейной части приращения (п. IV.8), то из (21) можно написать

$$dA = F(s)\,ds. \qquad (22)$$

Интегрируя, получим

$$A_{\text{общ}} = A(b) - A(a) = \int dA = \int_a^b F(s)\,ds;$$

часто пишут просто $A=\int F\,ds$. Хотя здесь пределы интегрирования не выписаны, но, конечно, это интеграл определенный, пределы подразумеваются.

На практике обычно вместо этого детального рассуждения пользуются следующим, более кратким: на протяжении бесконечно малого интервала пути ds силу можно считать постоянной, т. е. для соответствующего бесконечно малого приращения работы сразу получается формула (22), которую и интегрируют. Если подробно разобраться

в этом верном, но кратком рассуждении, то и получится приведенное выше детальное рассуждение. Мы вернемся еще к этому вопросу в п. XVI.4.

7. Дифференциальные уравнения с разделяющимися переменными. Соотношение (22) в общем виде можно записать так:

$$dy = f(x)\,dx, \tag{23}$$

где x и y — какие-то переменные, функционально связанные между собой. Интегрируя, получим

$$y_1 - y_0 = \int_{x_0}^{x_1} f(x)\,dx, \quad \text{где} \quad y_0 = y(x_0), \quad \text{а} \quad y_1 = y(x_1).$$

Уравнение (23) — простейшее дифференциальное уравнение. Более подробно дифференциальные уравнения будут изучаться в гл. XV, однако простейшие примеры, не требующие особой теории, можно показать уже сейчас. Так, встречаются уравнения вида

$$dy = \varphi(y)\,dx. \tag{24}$$

Такое уравнение нельзя сразу интегрировать, так как тогда в правой части под знаком интеграла будет стоять неизвестная функция $y(x)$. Требуется предварительно перенести y к dy, т. е. написать $\frac{dy}{\varphi(y)} = dx$, откуда

$$\int_{y_0}^{y_1} \frac{dy}{\varphi(y)} = x_1 - x_0.$$

Аналогичным образом уравнение

$$dy = f(x)\,\varphi(y)\,dx \tag{25}$$

интегрируется так:

$$\frac{dy}{\varphi(y)} = f(x)\,dx, \qquad \int_{y_0}^{y_1} \frac{dy}{\varphi(y)} = \int_{x_0}^{x_1} f(x)\,dx.$$

Уравнения (23)—(25) называются *дифференциальными уравнениями с разделяющимися переменными*, так как в результате простейших алгебраических преобразований в них удается отделить член с dx и x от члена с dy и y, после чего интегрирование немедленно выполняется.

Рассмотрим, например, задачу об истечении жидкости из цилиндрического сосуда, в дне которого проделано отверстие площади σ (рис. 251). Высота h уровня жидкости зависит от времени t, т. е. $h = h(t)$. Если жидкость невязкая и силами поверхностного натяжения можно пренебречь, то

скорость v истечения жидкости из сосуда с достаточной точностью описывается *законом Торичелли* (Э. Торичелли, 1608—1647, выдающийся итальянский физик и математик):

$$v=\sqrt{2gh}. \qquad (26)$$

На основании этого закона легко составить дифференциальное уравнение задачи. Будем рассуждать кратко, как в последнем абзаце п. 6. За время dt скорость истечения можно считать постоянной; значит, в силу формулы (26) вытекший объем равен $dV=\sigma\cdot v\,dt=\sigma\sqrt{2gh}\,dt$.

С другой стороны, тот же объем равен $dV=S\,|dh|=-S\,dh$; надо учесть, что h убывает и потому $dh<0$. Приравнивая оба выражения, получим уравнение

$$-S\,dh=\sigma\sqrt{2gh}\,dt \qquad (27)$$

Рис. 251.

типа (24). Для его интегрирования отнесем h к dh:

$$-\frac{S\,dh}{\sigma\sqrt{2gh}}=dt; \quad \int\limits_H^0 -\frac{S\,dh}{\sigma\sqrt{2gh}}=T,$$

где T — полное время истечения жидкости. Вычисляя, получим

$$-\frac{S}{\sigma\sqrt{2g}}2\sqrt{h}\,\Big|_{h=H}^{0}=T, \quad T=\frac{S}{\sigma}\sqrt{\frac{2H}{g}}.$$

8. Вычисление площадей плоских фигур. Применение определенного интеграла к вычислению площади криволинейной трапеции было рассмотрено в п. 2, а правило знаков показано на рис. 245.

Если надо найти всю заштрихованную на рис. 245 площадь в «арифметическом», а не «алгебраическом» смысле, то можно пользоваться формулой

$$S_1+S_2+S_3=\int\limits_a^b |f(x)|\,dx;$$

последний интеграл вычисляется, как описано в примере 1 п. 4.

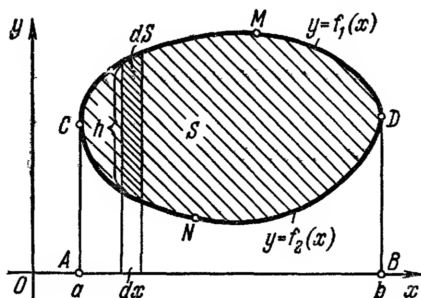


Рис. 252.

Вычисление площадей фигур, отличных от криволинейных трапеций, также осуществляется с помощью интегралов. Так, фигуру, изображенную на рис. 252, можно получить как разность двух криволинейных трапеций $ACMDBA$ и $ACNDBA$, т. е.

$$S=\int\limits_a^b f_1(x)\,dx-\int\limits_a^b f_2(x)\,dx=\int\limits_a^b [f_1(x)-f_2(x)]\,dx=\int\limits_a^b h(x)\,dx, \qquad (28)$$

где $h(x)$ — высота сечения фигуры прямой, параллельной оси y, на данной абсциссе x.

Формулу (28) можно понять также следующим образом. Если рассмотреть часть фигуры, лежащую левее прямой $x=\mathrm{const}$, то площадь этой части зависит от x. Если x получит бесконечно малое приращение dx (см. рис. 252), то к площади прибавится полоска, которую с точностью до малых высшего порядка можно считать прямоугольником (ср. вывод формул (21) и (22)). Отсюда $dS=h(x)\,dx$ и, интегрируя, получаем формулу (28).

Часто контур фигуры бывает задан *в параметрическом виде*. Тогда в рассматриваемых интегралах надо сделать замену переменных, приняв параметр за новую переменную интегрирования.

Вычислим, например, площадь, расположенную под одной из арок циклоиды (п. II.6), имеющей параметрические уравнения (II.12); при этом, чтобы получилась одна арка, должно быть $0\leqslant\psi\leqslant 2\pi$:

$$S=\int\limits_0^{2\pi R} y\,dx=\int\limits_{\psi=0}^{2\pi} R(1-\cos\psi)\,d\,[R(\psi-\sin\psi)]=R^2\int\limits_0^{2\pi}(1-\cos\psi)^2\,d\psi=3\pi R^2,$$

так как

$$\int(1-\cos\psi)^2\,d\psi=\int(1-2\cos\psi+\cos^2\psi)\,d\psi=\psi-2\sin\psi+\int\frac{1+\cos 2\psi}{2}\,d\psi=$$
$$=\frac{3}{2}\psi-2\sin\psi+\frac{\sin 2\psi}{4}+C.$$

Рассмотрим теперь площадь фигуры, ограниченной замкнутым контуром, заданным в параметрическом виде, $x=x(t)$, $y=y(t)$. Пусть при изменении t от α до γ контур (L) проходится один раз в положительном направлении, т. е. против часовой стрелки (рис. 253). Тогда

$$S=\int\limits_a^b y_1\,dx-\int\limits_a^b y_2\,dx.$$

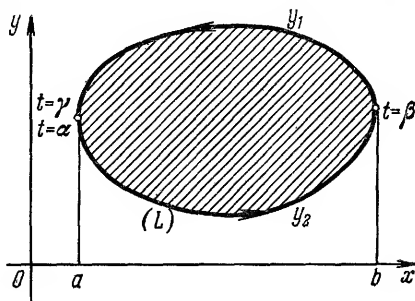


Рис. 253.

Однако первый интеграл равен $\int\limits_\gamma^\beta y\dot{x}\,dt$, так как когда t меняется от γ до β, то x меняется от a до b, а $y=y_1$ и $\dot{x}\,dt=dx$. Аналогично преобразуется второй интеграл, и мы получаем

$$S=\int\limits_\gamma^\beta y\dot{x}\,dt-\int\limits_\alpha^\beta y\dot{x}\,dt=-\int\limits_\alpha^\beta y\dot{x}\,dt-\int\limits_\beta^\gamma y\dot{x}\,dt=-\int\limits_\alpha^\gamma y\dot{x}\,dt. \qquad (29)$$

В силу свойства 10 п. 4 в окончательной формуле (29) значения $t=\alpha$ и $t=\gamma$ не обязаны отвечать крайней левой точке контура; важно только, чтобы он обходился ровно один раз.

Аналогично, проектируя контур на ось y, можно вывести формулу

$$S=\int_{\alpha}^{\gamma} x\dot{y}\,dt. \tag{30}$$

Если же сложить формулы (29) и (30), то получится

$$S=\frac{1}{2}\int_{\alpha}^{\gamma}(x\dot{y}-y\dot{x})\,dt. \tag{31}$$

Если бы с возрастанием t контур проходился в отрицательном направлении, то во всех формулах (29)—(31) надо было бы переменить знак

Например, площадь эллипса с параметрическими уравнениями (11. 26) на основе формулы (31) равна

$$S=\frac{1}{2}\int_{0}^{2\pi}[a\cos t\cdot b\cos t-b\sin t\,(-a\sin t)]\,dt=\frac{1}{2}\int_{0}^{2\pi}ab\,dt=\pi ab.$$

Перейдем к выражению площади в полярных координатах. Пусть линия задана полярным уравнением $\rho=f(\varphi)$ и мы хотим вычислить

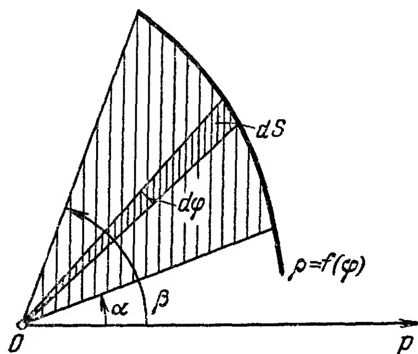


Рис 254.

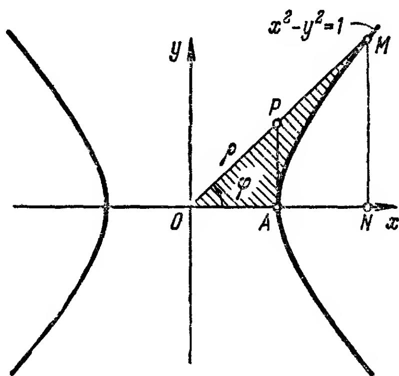


Рис 255.

площадь «сектора» $\alpha\leqslant\varphi\leqslant\beta$ (рис. 254). Если угол φ увеличивается на $d\varphi$, то к площади прибавляется кусочек, который с точностью до малых высшего порядка можно принять за равнобедренный треугольник с высотой ρ и основанием $\rho\,d\varphi$ (почему?). Значит,

$$dS=\frac{1}{2}\rho\rho\,d\varphi,\qquad S=\frac{1}{2}\int_{\alpha}^{\beta}\rho^2\,d\varphi. \tag{32}$$

В качестве примера найдем площадь, заштрихованную на рис 255. Переход в уравнении гиперболы к полярным координатам дает

$$\rho^2\cos^2\varphi-\rho^2\sin^2\varphi=1,\quad \text{т. е.}\quad \rho^2=\frac{1}{\cos^2\varphi-\sin^2\varphi},$$

и по формуле (32) получаем (проверьте!)

$$S=\frac{1}{2}\int_0^{\varphi}\frac{1}{\cos^2\varphi-\sin^2\varphi}\,d\varphi=\frac{1}{4}\ln\frac{1+\operatorname{tg}\varphi}{1-\operatorname{tg}\varphi}.$$

Отсюда получается интересное следствие Так как

$$\frac{1+\operatorname{tg}\varphi}{1-\operatorname{tg}\varphi}=e^{4S},\quad \operatorname{tg}\varphi=\frac{e^{4S}-1}{e^{4S}+1}=\frac{e^{2S}-e^{-2S}}{e^{2S}+e^{-2S}}=\operatorname{th}2S$$

(п. I.28), то

$$NM=\rho\sin\varphi=\frac{\sin\varphi}{\sqrt{\cos^2\varphi-\sin^2\varphi}}=\frac{\operatorname{tg}\varphi}{\sqrt{1-\operatorname{tg}^2\varphi}}=\frac{\operatorname{th}2S}{\sqrt{1-\operatorname{th}^2 2S}}=\frac{\operatorname{th}2S}{\frac{1}{\operatorname{ch}2S}}=\operatorname{sh}2S.$$

Аналогично находим, что $ON=\rho\cos\varphi=\operatorname{ch}2S$, $AP=\operatorname{tg}\varphi=\operatorname{th}2S$ Сравнение этих результатов с рис. 256, где $\varphi=2S$, $MN=\sin 2S$, $ON=\cos 2S$, $AP=\operatorname{tg}2S$, показывает геометрическую причину связи тригонометрических (круговых) функций с гиперболическими и раскрывает происхождение термина «гиперболические» синус, косинус и тангенс

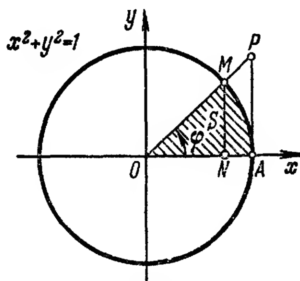


Рис 256

9. Длина дуги. Дифференциал дуги уже встречался в нашем курсе (п. VII.23); будем обозначать его dL:

$$dL=\sqrt{dx^2+dy^2+dz^2}=$$
$$=\sqrt{\dot x^2+\dot y^2+\dot z^2}\,dt.$$

При этом мы перед корнем берем только $+$, считая $dL>0$. Отсюда если дуга ограничена значениями $t=\alpha$ и $t=\beta$, то ее длина

$$L=\int_\alpha^\beta\sqrt{\dot x^2+\dot y^2+\dot z^2}\,dt.$$

Формула для длины плоской линии соответственно упрощается. Если же линия задана уравнением вида $y=f(x)$, то при $a\leqslant x\leqslant b$

$$L=\int_\alpha^\beta\sqrt{dx^2+dy^2}=\int_a^b\sqrt{1+y'^2}\,dx.$$

Например, длина циклоиды между ее соседними точками возврата в силу уравнений (II 12) и симметрии циклоиды равна

$$L=2\int_0^\pi R\sqrt{(\psi-\sin\psi)'^2+(1-\cos\psi)'^2}\,d\psi=2R\int_0^\pi\sqrt{(1-\cos\psi)^2+\sin^2\psi}\,d\psi=$$
$$=2R\int_0^\pi\sqrt{2-2\cos\psi}\,d\psi=4R\int_0^\pi\sin\frac{\psi}{2}\,d\psi=-8R\cos\frac{\psi}{2}\Big|_0^\pi=8R.$$

Какой простой результат!

Дифференциал дуги в полярных координатах легко получается из рис. 257:

$$dL=\sqrt{(d\rho)^2+(\rho\,d\varphi)^2}. \qquad (33)$$

Отсюда если уравнение линии дано в полярных координатах в виде $\rho=f(\varphi)$, то ее длина при $\alpha\leqslant\varphi\leqslant\beta$ равна

$$L=\int\limits_{\varphi=\alpha}^{\beta}\sqrt{d\rho^2+\rho^2\,d\varphi^2}=\int\limits_{\alpha}^{\beta}\sqrt{\left(\frac{d\rho}{d\varphi}\right)^2+\rho^2}\,d\varphi.$$

Рис. 257.

Проверьте, что выражение (33) можно получить также из формул

$$dL=\sqrt{dx^2+dy^2},\quad x=\rho\cos\varphi,\quad y=\rho\sin\varphi.$$

Например, пользуясь полярным уравнением кардиоиды, указанным на рис. 72, найдем ее длину

$$L=2\int\limits_0^{\pi}\sqrt{4a^2\sin^2\varphi+4a^2(1-\cos\varphi)^2}\,d\varphi=16a$$

(проверьте!).

10. Объем тела. Пусть требуется найти объем тела, если известны площади сечений плоскостями, перпендикулярными к некоторой оси, которую мы примем за ось x (рис. 258), $S=S(x)$. Если отсчитывать объем от левого конца тела до указанной плоскости, то, когда x увеличится на $\Delta x=dx$ и эта плоскость передвинется направо, к объему прибавится «ломоть», который с точностью до малых высшего порядка можно считать цилиндром с широким основанием и маленькой высотой. Отсюда

$$\Delta V=S(x)\,\Delta x+\text{малые высшего порядка},$$

т. е.

$$dV=S(x)\,dx,$$

и если x меняется от a до b,

$$V=\int\limits_a^b S(x)\,dx. \qquad (34)$$

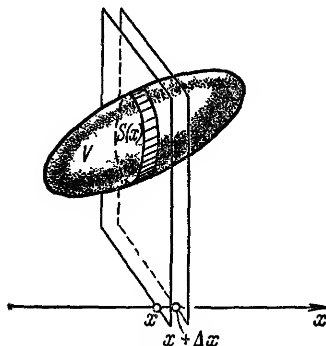


Рис. 258.

В качестве примера найдем объем *цилиндрического копыта*, отсекаемого от прямого кругового цилиндра плоскостью, проходящей через диаметр основания (рис 259). Из подобия треугольников ABC и $A'B'C'$ получаем, что площадь заштрихованного сечения равна

$$S=\frac{1}{2}RH\,\frac{R^2-x^2}{R^2}=\frac{R^2-x^2}{2R}H$$

и по формуле (34)

$$V=2\int_0^R S\,dx=2\int_0^R \frac{R^2-x^2}{2R}H\,dx=\frac{H}{R}\left(R^2x-\frac{x^3}{3}\right)\Big|_0^R=\frac{2}{3}R^2H.$$

Интересно, что в ответ не входит π, как это можно было бы ожидать.

Рассмотрим, в частности, *объем тела вращения*. Пусть линия с уравнением $y=f(x)$ вращается в пространстве вокруг оси x;

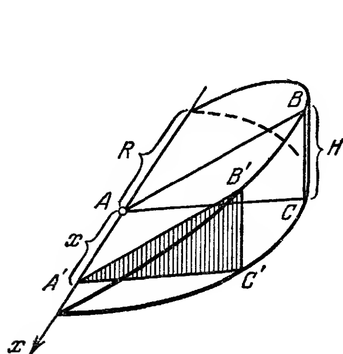


Рис. 259.

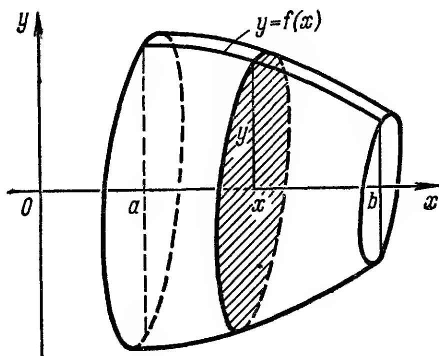


Рис. 260.

тогда площадь поперечного сечения равна $S=\pi y^2$ (см. рис. 260) и по формуле (34)

$$V=\pi\int_a^b y^2\,dx. \tag{35}$$

Например, шар получается в результате вращения полуокружности с уравнением $y=\sqrt{R^2-x^2}$, поэтому объем шара

$$V=\pi\cdot 2\int_0^R(\sqrt{R^2-x^2})^2\,dx=2\pi\left(R^2x-\frac{x^3}{3}\right)\Big|_0^R=\frac{4}{3}\pi R^3.$$

Интересно сравнить этот вывод с длинным и искусственным выводом объема шара в курсе элементарной математики.

11. Площадь поверхности вращения. Формула для площади поверхности произвольного вида будет выведена в п. XVI.10. Однако вычисление площади поверхности вращения можно показать сейчас. Пусть линия $y=f(x)>0$ вращается вокруг оси x (рис. 261); будем проводить плоскости, перпендикулярные к оси вращения, и рассматривать площадь поверхности, расположенной левее любого такого сечения.

Если плоскость переместится на dx, то к площади добавится «элементарное кольцо», заштрихованное на рис. 261. Разрезав это

кольцо и развернув его, получим полоску ширины dL и длины $2\pi y$, так как y — радиус кольца. Отсюда

$$dS = 2\pi y\, dL = 2\pi y \sqrt{1+y'^2}\, dx$$

и тем самым

$$S = 2\pi \int_{x=a}^{b} y\, dL = 2\pi \int_{a}^{b} y \sqrt{1+y'^2}\, dx. \qquad (36)$$

Вычислим, например, площадь части параболоида вращения, отсеченной плоскостью, перпендикулярной к оси вращения (рис. 262), если даны радиус

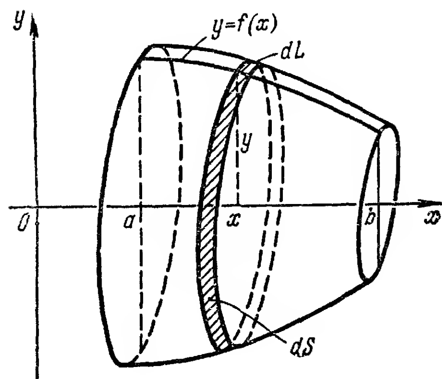


Рис. 261.

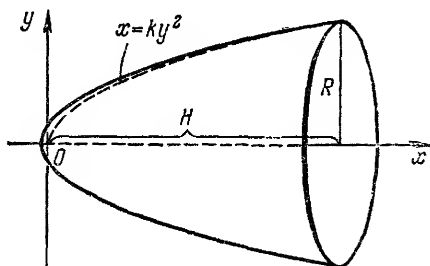


Рис. 262.

основания R и высота H. Так как вращающейся линией служит парабола с осью по оси x, то уравнение линии имеет вид $x=ky^2$. Константу k надо подобрать так, чтобы парабола прошла через точку (H, R), т. е. $H=kR^2$, $k=\frac{H}{R^2}$, и окончательно уравнение линии будет

$$x=\frac{H}{R^2}y^2, \quad \text{т. е.} \quad y=R\sqrt{\frac{x}{H}}.$$

Пользуясь формулой (36), получим

$$S=2\pi\int_0^H R\sqrt{\frac{x}{H}}\sqrt{1+\left[\left(R\sqrt{\frac{x}{H}}\right)'\right]^2}\,dx=$$

$$=2\pi\frac{R}{\sqrt{H}}\int_0^H \sqrt{x}\sqrt{1+\frac{R^2}{H}\cdot\frac{1}{4x}}\,dx=\pi\frac{R}{H}\int_0^H\sqrt{4xH+R^2}\,dx=$$

$$=\pi\frac{R}{H}\frac{(4xH+R^2)^{3/2}}{\left(\frac{3}{2}\right)\cdot 4H}\Bigg|_0^H=\frac{\pi R}{6H^2}\left[(4H^2+R^2)^{3/2}-R^3\right].$$

§ 3. Численное интегрирование

12. Общие замечания. Описанный в п. 3 основной способ вычисления определенного интеграла с помощью неопределенного на практике не всегда возможен и целесообразен. Как было указано в п. XIII.11, многие неопределенные интегралы даже от элементарных функций не выражаются через элементарные функции или имеют чрезмерно громоздкие выражения. Кроме того, функция, которую надо проинтегрировать, может быть задана не формулой, а как-либо иначе. В этих случаях для вычисления интегралов применяется целый ряд способов, о которых мы сейчас дадим общее представление.

1. *Интегралы могут выражаться через хорошо изученные и затабулированные неэлементарные «специальные» функции.*

Такими являются, например, *функция ошибок*

$$\operatorname{Erf} x = \int_0^x e^{-t^2}\,dt \qquad (-\infty < x < \infty),$$

от английского «error function», *интегралы Френеля* (О. Френе́ль, 1788—1827, выдающийся французский физик, создатель волновой теории света)

$$C(x) = \int_0^x \cos\frac{\pi t^2}{2}\,dt, \qquad S(x) = \int_0^x \sin\frac{\pi t^2}{2}\,dt \qquad (-\infty < x < \infty);$$

интегральная показательная функция

$$\operatorname{Ei} x = \int_{-\infty}^x \frac{e^t}{t}\,dt \qquad (-\infty < x < 0);$$

интегральный синус и *интегральный косинус*

$$\operatorname{Si} x = \int_0^x \frac{\sin t}{t}\,dt \qquad (-\infty < x < \infty),$$

$$\operatorname{Ci} x = \int_\infty^x \frac{\cos t}{t}\,dt \qquad (0 < x < \infty)$$

и многие другие функции. Отметим, что интегралы с бесконечными пределами будут рассмотрены в § 4.

Например, чтобы вычислить интеграл $I = \int_0^1 \frac{\sin^2 x}{x^2}\,dx$, произведем интегрирование по частям: $u = \sin^2 x$, $dv = x^{-2}\,dx$,

$$I = -\frac{\sin^2 x}{x}\Big|_0^1 + \int_0^1 \frac{2\sin x\cos x}{x}\,dx = -\sin^2 1 + \int_0^1 \frac{\sin 2x}{x}\,dx = |\,2x = t\,| =$$

$$= -\sin^2 1 + \int_0^2 \frac{\sin t}{t}\,dt = -\sin^2 1 + \operatorname{Si} 2 = 0{,}8973;$$

значение интегрального синуса взято из книги [53]. В ней описано большое число специальных функций

2. *Иногда удается найти точное значение для определенного интеграла с теми или иными пределами, не вычисляя неопределенного интеграла.* Это часто бывает довольно трудно, хотя некоторые примеры будут даны в дальнейшем (см., например, формулу (72)). Многие из таких интегралов собраны в книге [13].

Например, в этой книге мы можем прочитать, что

$$\int_0^{\frac{\pi}{2}} \operatorname{tg}^p x\,dx = \frac{\pi}{2}\left(\cos\frac{p\pi}{2}\right)^{-1} \quad (-1 < p < 1), \qquad \int_0^{\pi} \ln\sin x\,dx = -\pi\ln 2 \text{ и т. п.},$$

хотя соответствующие неопределенные интегралы не являются элементарными функциями.

3. *Часто применяются разложения подынтегральной функции в ряды различного вида.* Подробное описание этого метода будет дано в гл. XVII, однако простые степенны́е ряды, о которых говорилось в п. IV.16, можно применять уже сейчас.

Например, применяя ряд (IV.55) для e^x, получим

$$\int_0^1 \frac{e^x-1}{x}\,dx = \int_0^1 \frac{1+\frac{x}{1!}+\frac{x^2}{2!}+\ldots-1}{x}\,dx =$$

$$= \int_0^1 \left(\frac{1}{1!}+\frac{x}{2!}+\frac{x^2}{3!}+\ldots\right)dx = \frac{1}{1!}+\frac{1}{2\cdot 2!}+\frac{1}{3\cdot 3!}+\ldots = 1{,}318$$ (с точностью до 0,001).

Как говорилось в п. IV.16, к подобным рядам на практике надо относиться как к конечным суммам, число членов которых берется в соответствии с требуемой степенью точности.

4. *Графическое интегрирование* применяется, если подынтегральная функция задана графиком. Оно основано на геометрическом смысле определенного интеграла (п. 2), равного площади соответствующей криволинейной трапеции. Эту площадь можно подсчитать примитивно, изобразив график на миллиметровке и считая клеточки, либо же с помощью специального инструмента — *планиме́тра*. Площадь фигуры произвольного вида считывается с циферблата планиметра после обвода ее контура штырем планиметра. Поскольку планиметр представляет собой простой механизм, то интегрирование с его помощью называется также *механическим интегрированием*.

5. Наиболее универсальными методами, пригодными к интегралам от произвольных функций, заданных любым способом, в особенности таблично (это, в частности, удобно при применении вычислительных машин), являются методы численного интегрирования (п. 13).

13. Формулы численного интегрирования. Эти формулы дают приближенные значения определенного интеграла, если известны значения подынтегральной функции в некоторых точках (*узлах*) интервала интегрирования.

Начнем с наиболее простой формулы. Пусть требуется вычислить интеграл

$$\int_a^b y\,dx, \qquad y=f(x), \tag{37}$$

причем предполагается, что интервал интегрирования $a \leqslant x \leqslant b$ разбит на некоторое число n равных частей и заданы или вычислены значения подынтегральной функции в точках деления.

Обозначим

$$\frac{b-a}{n}=h, \qquad f(a)=y_0, \qquad f(a+h)=y_1,$$
$$f(a+2h)=y_2, \quad \ldots, \quad f(a+nh)=f(b)=y_n.$$

Если провести ординаты для каждого из полученных узлов, то криволинейная трапеция, площадь которой равна интегралу (37), разбивается (рис. 263) на n столбиков, каждый из которых также представляет собой криволинейную трапецию. Заменим эти столбики прямолинейными («школьными») трапециями, построенными на крайних ординатах, как показано на рис. 263. Площади этих трапеций последовательно равны

$$\frac{y_0+y_1}{2}h, \qquad \frac{y_1+y_2}{2}h, \quad \ldots, \quad \frac{y_{n-1}+y_n}{2}h.$$

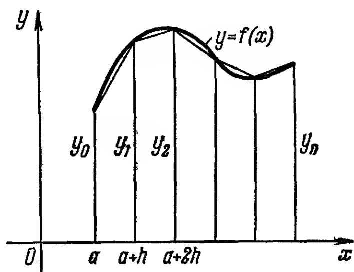


Рис. 263.

Сложив эти площади, мы получим площадь многоугольной фигуры, вписанной в исходную криволинейную трапецию. Если n достаточно велико, т. е. h достаточно мало́, то эта площадь приближенно равна площади криволинейной трапеции, т. е. интегралу, и мы получаем

$$\int_a^b y\,dx \approx \frac{y_0+y_1}{2}h+\frac{y_1+y_2}{2}h+\ldots+\frac{y_{n-1}+y_n}{2}h$$

или

$$\int_a^b y\,dx \approx h\left(\frac{y_0+y_n}{2}+y_1+y_2+\ldots+y_{n-1}\right). \tag{38}$$

Это — так называемая *формула трапеций*.

Формуле трапеций можно дать истолкование, не зависящее от ее геометрического смысла. По сути дела, мы перед интегрированием заменили на каждом из интервалов $a \leqslant x \leqslant a+h$, $a+h \leqslant x \leqslant a+2h$

и т. д. рассматриваемую функцию $y=f(x)$ на линейную, принимающую те же значения в концевых точках, т. е. произвели линейную интерполяцию (п. I.22), после чего уже осуществили интегрирование. Но в § V.2 были приведены интерполяционные формулы, дающие приближение функции с гораздо большей точностью, чем при линейной интерполяции. Поэтому формулы численного интегрирования, полученные на основе этих интерполяционных формул, являются значительно более точными, чем формула (38).

Если применить интерполяционный многочлен второй степени, то получится *формула Симпсона*, названная по имени английского математика Т. Си́мпсона (1710—1761), который ее вывел. Предположим сначала, что значения интегрируемой функции даны в трех точках

$$y(x_0)=y_0, \qquad y(x_0+h)=y_1, \qquad y(x_0+2h)=y_2.$$

Тогда интерполяционный многочлен второй степени, принимающий в этих точках такие же значения, выписывается по формуле Ньютона (см. п. V.8):

$$P(x)=y_0+\Delta y_0\frac{s}{h}+\frac{\Delta^2 y_0}{2}\frac{s}{h}\left(\frac{s}{h}-1\right) \qquad (s=x-x_0).$$

Отсюда для интеграла получаем выражение

$$\int\limits_{x_0}^{x_0+2h} P(x)\,dx=\left|\begin{array}{l} x-x_0=s\\ dx=ds\\ 0\leqslant s\leqslant 2h\end{array}\right|=\int\limits_0^{2h}\left[y_0+\Delta y\frac{s}{h}+\frac{\Delta^2 y_0}{2}\left(\frac{s^2}{h^2}-\frac{s}{h}\right)\right]ds=$$
$$=y_0\cdot 2h+\Delta y_0\cdot 2h+\frac{\Delta^2 y_0}{2}\cdot\frac{2}{3}h.$$

Если, далее, подставить

$$\Delta y_0=y_1-y_0,$$
$$\Delta^2 y_0=\Delta y_1-\Delta y_0=(y_2-y_1)-(y_1-y_0)=y_2-2y_1+y_0,$$

то после приведения подобных членов получим

$$\int\limits_{x_0}^{x_0+2h} f(x)\,dx\approx\int\limits_{x_0}^{x_0+2h} P(x)\,dx=2h\left[y_0+(y_1-y_0)+\right.$$
$$\left.+\frac{1}{6}(y_2-2y_1+y_0)\right]=h\,\frac{y_0+4y_1+y_2}{3}. \qquad (39)$$

Если теперь отрезок интегрирования $a\leqslant x\leqslant b$ разбит на $2n$ равных частей с помощью точек деления

$$x_0=a, \quad x_1=a+h, \quad x_2=a+2h, \quad \ldots, \quad x_{2n}=a+2nh=b$$
$$\left(h=\frac{b-a}{2n}\right),$$

то к каждой паре из этих частей можно применить формулу (39)

$$\int_{x_0}^{x_0+2h} y\,dx \approx h\frac{y_0+4y_1+y_2}{3},$$

$$\int_{x_0+2h}^{x_0+4h} y\,dx \approx h\frac{y_2+4y_3+y_4}{3}, \ \ldots, \ \int_{x_0+(2n-2)h}^{x_0+2nh} y\,dx \approx h\frac{y_{2n-2}+4y_{2n-1}+y_{2n}}{3}.$$

Сложив эти формулы и приведя подобные члены, получим формулу Симпсона:

$$\int_a^b y\,dx \approx \frac{h}{3}[(y_0+y_{2n})+2(y_2+y_4+\ldots+y_{2n-2})+$$
$$+4(y_1+y_3+\ldots+y_{2n-1})]. \quad (40)$$

Перейдем теперь к оценке точности формул (38) и (40). Из формулы Ньютона (V.27) видно, что при осуществлении линейной интерполяции получается ошибка порядка $\Delta^2 y$, т. е порядка h^2 (п. V.7). При интегрировании она множится на длину h интервала интегрирования (см. формулу (18)), т. е. приобретает порядок h^3. Но формула (38) получается в результате сложения n приближенных формул, каждая с ошибкой порядка h^3, поэтому суммарная ошибка имеет порядок

$$n\cdot h^3 = \frac{b-a}{h}h^3 = (b-a)\cdot h^2,$$

т. е. порядок h^2. Например, если число точек деления увеличить в два раза, то точность формулы (38) улучшится примерно в четыре раза.

Если рассуждать так же о формуле (40), то может показаться, что ее ошибка имеет порядок h^3. На самом же деле точность выше. Действительно, когда мы в формуле Ньютона отбрасываем член с $\Delta^3 y_0$, то делаем ошибку порядка h^3; но интеграл от этого члена равен нулю

$$\int_{x_0}^{x_0+2h} \frac{s}{h}\left(\frac{s}{h}-1\right)\left(\frac{s}{h}-2\right)dx = \int_0^{2h}\left[\left(\frac{s}{h}\right)^3-3\left(\frac{s}{h}\right)^2+2\frac{s}{h}\right]ds =$$
$$= \left(\frac{s^4}{4h^3}-\frac{3s^3}{3h^2}+\frac{2s^2}{2h}\right)\Big|_0^{2h} = 0.$$

Поэтому после интегрирования ошибку дает только следующий член формулы Ньютона, имеющий порядок h^4. Значит, ошибка в формуле (39) имеет порядок h^5, а в окончательной формуле (40)—порядок h^4; например, если увеличить число точек деления в два раза, то точность формулы (40) улучшится в 16 раз. В то же время формула (40) для применения ненамного сложнее формулы (38).

Рассмотрим для примера интеграл

$$I = \int_0^1 \frac{1}{1+x^2}dx = \operatorname{arctg} x\Big|_0^1 = \frac{\pi}{4} = 0{,}785.$$

Если бы мы не знали ответа, то интеграл можно было бы приближенно вычислить по формулам (38) или (40). Примем для простоты $n=2$, т. е.

$$h=0{,}5;\quad x_0=0;\quad x_1=0{,}5;\quad x_2=1;\quad y_0=1{,}000;\quad y_1=0{,}800;\quad y_2=0{,}500.$$

Подсчет по формуле (38) дает

$$I \approx 0{,}5\left(\frac{1{,}000+0{,}500}{2}+0{,}800\right)=0{,}775; \quad \text{ошибка } \delta_{(38)} \approx 1{,}3\%.$$

Подсчет по формуле (40) дает

$$I \approx \frac{0{,}5}{3}(1{,}000+0{,}500+4\cdot 0{,}800)=0{,}783; \text{ ошибка } \delta_{(40)} \approx 0{,}3\%.$$

Если бы мы приняли $n=10$, то получилось бы $\delta_{(38)} \approx 0{,}05\%$; $\delta_{(40)} \approx 10^{-6}\%$.

В книгах, посвященных численным метода математики, рассматриваются еще более точные формулы, чем формула Симпсона. Отметим, что иногда применяются формулы с *неравноотстоящими узлами*.

§ 4. Несобственные интегралы

До сих пор при рассмотрении определенных интегралов мы считали, что интервал интегрирования конечен и что подынтегральная функция на нем не обращается в бесконечность; такие интегралы мы будем называть *интегралами в собственном смысле слова* или, коротко, *собственными*. Если хотя бы одно из этих двух условий не выполнено, то интеграл называется *несобственным*. Собственный интеграл (см. п. 2) всегда имеет определенное численное значение. В отличие от этого несобственные интегралы, к которым мы сейчас переходим, не все имеют такое значение.

14. Интеграл с бесконечным пределом интегрирования. Рассмотрим сначала интеграл вида

$$I=\int_a^\infty f(x)\,dx, \tag{41}$$

где нижний предел a и подынтегральная функция $f(x)$ при $a \leqslant x < \infty$ предполагаются конечными. Такой интеграл является несобственным из-за того, что его верхний предел бесконечный.

Чтобы придать точный смысл интегралу (41), пользуются тем же приемом, с помощью которого в п. III.6 было уточнено понятие суммы бесконечного ряда. Именно, сначала «отрезают» бесконечность, т. е. рассматривают интеграл

$$\int_a^N f(x)\,dx, \tag{42}$$

где N — большое, но конечное число. Интеграл (42) собственный и имеет определенное численное значение. Затем принимают, что N стремится к бесконечности, т. е. к тому, чему оно должно равняться для интеграла (41). При $N\to\infty$ интеграл (42) как-то меняется; *если он имеет при этом определенный конечный предел, то интеграл* (41)

называют сходящимся и полагают

$$\int_a^\infty f(x)\,dx = \lim_{N\to\infty}\int_a^N f(x)\,dx. \qquad (43)$$

Если конечного предела нет, то интеграл (41) *называется расходящимся; в этом случае определенного численного значения мы ему приписывать не будем*, во всяком случае в этой книге. Итак, *мы принимаем, что несобственный интеграл имеет определенное численное значение только в случае его сходимости.*

Отметим частный случай расходимости: если интеграл (42) при $N\to\infty$ имеет бесконечный предел, то интеграл (41) называется расходящимся к бесконечности; в этом случае можно пользоваться формулой (43).

Рассмотрим несколько примеров. Пусть на точку T действует сила, направленная к некоторой фиксированной точке O и по модулю обратно пропорциональная квадрату расстояния от O; так ведут себя, в частности, гравитационная и электростатическая силы. Пусть при этом требуется найти работу, которую надо затратить, чтобы удалить точку T из некоторого положения T_0 на бесконечность; эта работа называется *потенциалом* рассматриваемой силы.

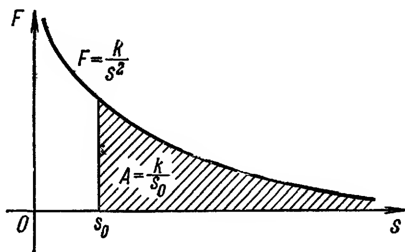


Рис. 264.

Для ее подсчета запишем силу

$$F=\frac{k}{s^2}\ (s=OT),$$

где k — некоторый коэффициент пропорциональности, и работу

$$A=\int_{s_0}^\infty \frac{k}{s^2}\,ds \qquad (44)$$

(см п. 6). Это — несобственный интеграл, и по формуле (43)

$$A=\lim_{N\to\infty}\int_{s_0}^N \frac{k}{s^2}\,ds=\lim_{N\to\infty}-\frac{k}{s}\Big|_{s=s_0}^N=\lim_{N\to\infty}\left(\frac{k}{s_0}-\frac{k}{N}\right)=\frac{k}{s_0}.$$

Итак, интеграл (44) сходится, и мы получаем, что рассматриваемый потенциал обратно пропорционален первой степени расстояния от точки O. Сначала может показаться странным, что при удалении на бесконечное расстояние получается конечная работа, хотя сила теоретически никогда не прекращает своего действия. Но дело в том, что эта сила при удалении так ослабевает, что затрачиваемая работа, хотя все время и нарастает, но стремится не к бесконечности, а к конечной величине. Геометрический смысл полученного результата показан на рис. 264: несмотря на то, что заштрихованная фигура простирается в бесконечность, ее высота убывает с такой скоростью, что общая ее площадь оказывается конечной.

Реально s изменяется не до бесконечности, так как все физические величины конечны, а до какого-то очень большого, но конечного значения S, т. е.

взамен (44) надо рассматривать интеграл

$$\int_{s_0}^{S} \frac{k}{s^2}\, ds \tag{45}$$

Однако для достаточно больших S этот интеграл практически не меняется, так что его можно заменить на «предельный» интеграл (44), который проще в теоретических исследованиях, тем более, что значение S точно не известно. Сходимость интеграла (44) и означает возможность замены «реального» интеграла (45) при больших S на интеграл (44), физически это значит, что на большом удалении от точки O ее действием можно практически пренебречь. При такой замене точно знать S и не нужно, важно только знать, что S велико

В качестве второго примера рассмотрим несобственный интеграл

$$\int_{1}^{\infty} \frac{1}{x}\, dx. \tag{46}$$

Так как

$$\lim_{N \to \infty} \int_{1}^{N} \frac{1}{x}\, dx = \lim_{N \to \infty} \ln x \Big|_1^N = \lim_{N \to \infty} \ln N = \ln \infty = \infty,$$

то интеграл (46) расходится к бесконечности и можно написать

$$\int_{1}^{\infty} \frac{1}{x}\, dx = \infty.$$

Рассмотрим, наконец, несобственный интеграл

$$\int_{0}^{\infty} \sin x\, dx. \tag{47}$$

В данном случае предел

$$\lim_{N \to \infty} \int_{0}^{N} \sin x\, dx = \lim_{N \to \infty} (-\cos x)\Big|_0^N = \lim_{N \to \infty} (1 - \cos N)$$

не существует ни конечный, ни бесконечный, так как $\cos N$ при $N \to \infty$ все время колеблется между $+1$ и -1 Значит интеграл (47) расходится, но не к бесконечности, а колеблясь между конечными значениями 0 и 2. Он не имеет ни конечного, ни бесконечного значений.

15. Основные свойства интегралов с бесконечным пределом интегрирования. На несобственные интегралы вида (41) непосредственно распространяются многие свойства собственных интегралов.

Прежде всего, если при $a \leqslant x < \infty$ известна первообразная функция $F(x)$ к $f(x)$, то

$$\int_{a}^{\infty} f(x)\, dx = \lim_{N \to \infty} \int_{a}^{N} f(x)\, dx = \lim_{N \to \infty} [F(N) - F(a)] = F(\infty) - F(a),$$

так как под $F(\infty)$ как раз и понимается $\lim\limits_{N\to\infty} F(N)$. Значит, несобственный интеграл (41) можно вычислять по той же основной формуле (11), что и собственные интегралы. При этом само выражение $F(\infty)$ покажет, сходится интеграл или расходится.

Так, в примерах (44), (46) и (47) можно было вычислять просто:

$$\int\limits_{s_0}^{\infty} \frac{k}{s^2}\,ds = -\frac{k}{s}\Bigg|_{s=s_0}^{\infty} = \frac{k}{s_0} - \frac{k}{\infty} = \frac{k}{s_0},$$

$$\int\limits_{1}^{\infty} \frac{1}{x}\,dx = \ln x\Bigg|_{1}^{\infty} = \ln\infty - \ln 1 = \infty,$$

$$\int\limits_{0}^{\infty} \sin x\,dx = -\cos x\Bigg|_{0}^{\infty} = -\cos\infty + 1 \text{ не существует,}$$

так как выражение $\cos\infty$ не имеет смысла.

Все основные свойства п. 4 сохраняются и для несобственных интегралов, конечно, если иметь в виду осложнения, которые получаются в случае расходимости интегралов. В то же время формулы (18) и (19) здесь уже не действуют, так как *интеграл от константы, не равной нулю, по бесконечному интервалу всегда расходится*. Отметим еще такое простое свойство: если интеграл (41) сходится, то при $N\to\infty$

$$\int\limits_{N}^{\infty} f(x)\,dx = \int\limits_{a}^{\infty} f(x)\,dx - \int\limits_{a}^{N} f(x)\,dx \to 0.$$

Если неопределенный интеграл взять затруднительно, то обычно начинают с выяснения его сходимости или расходимости на основе признаков, к которым мы сейчас переходим.

Заметим прежде всего, что факт сходимости или расходимости интеграла (47) зависит только от поведения функции $f(x)$ *на бесконечности*, т. е. при достаточно больших x. Другими словами, интегралы $\int\limits_{a}^{\infty} f(x)\,dx$ и $\int\limits_{b}^{\infty} f(x)\,dx$ сходятся или расходятся одновременно, если $f(x)$ между a и b не обращается в бесконечность. Действительно, эти интегралы отличаются на собственный интеграл, который имеет вполне определенное числовое значение и не может нарушить сходимость, если она была, и создать сходимость, если ее не было.

Рассмотрим сначала *интеграл от положительной функции*

$$\int\limits_{a}^{\infty} f(x)\,dx, \quad f(x) \geqslant 0. \tag{48}$$

Такой интеграл может либо сходиться, либо расходиться к бесконечности, так как интеграл, взятый от a до N, с ростом N растет, а возрастающая величина имеет конечный предел либо стремится к бесконечности (п. III.5). Эта сходимость или расходимость означает геометрически, что площадь бесконечной фигуры, заштрихованной на рис. 265, получится конечной или бесконечной. Факт сходимости или расходимости интеграла (48) можно записать так:

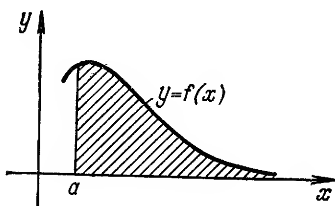


Рис. 265.

$$\int_a^{\infty} f(x)\,dx < \infty \text{ или соответственно } \int_a^{\infty} f(x)\,dx = \infty.$$

Конечно, такую запись нельзя применять по отношению к расходящимся колеблющимся интегралам типа (47).

Самым простым признаком сходимости является *признак сравнения: если*

$$0 \leqslant g(x) \leqslant f(x) \qquad (a \leqslant x < \infty) \tag{49}$$

и $\int_a^{\infty} f(x)\,dx < \infty$, *т. е. этот интеграл сходится, то и* $\int_a^{\infty} g(x)\,dx < \infty$, *т. е. и этот интеграл сходится.* Этот признак вытекает из интегрирования неравенства (49) или из геометрического смысла сходимости (см. рис. 265). Из этого признака следует, что если в условиях (49) интеграл от $g(x)$ расходится (равен бесконечности), то интеграл от $f(x)$ тоже расходится.

Применяется также следующий признак: *если*

$$\frac{f(x)}{g(x)} \underset{x\to\infty}{\longrightarrow} k \neq 0, \qquad k \neq \infty, \tag{50}$$

то интегралы

$$\int_a^{\infty} f(x)\,dx \quad \text{и} \quad \int_b^{\infty} g(x)\,dx$$

сходятся или расходятся одновременно (хотя в случае сходимости их значения могут существенно различаться, даже если $k=1$ и $a=b$). Действительно, из условия (50) вытекает, что на бесконечности обе функции, $f(x)$ и $g(x)$, не могут существенно превзойти одна другую, точнее говоря, $f(x) \sim kg(x)$ при $x \to \infty$, где $\sim$ — знак эквивалентности (см. п. III.7). Поэтому если для одной из них заштрихованная фигура на рис. 265 имеет конечную площадь, то и для другой тоже.

Чаще всего заданный интеграл вида (48) сравнивают с *интегралом от степенной функции*

$$I=\int_1^\infty \frac{1}{x^p}\,dx, \tag{51}$$

который легко исследовать непосредственно. При $p>1$ имеем

$$I=\int_1^\infty x^{-p}\,dx=\frac{x^{-p+1}}{-p+1}\Big|_1^\infty=\frac{1}{(-p+1)\,x^{p-1}}\Big|_1^\infty=\frac{1}{\infty}-\frac{1}{-p+1}=\frac{1}{p-1}<\infty,$$

тогда как при $p<1$

$$I=\frac{x^{-p+1}}{-p+1}\Big|_1^\infty=\frac{x^{1-p}}{1-p}\Big|_1^\infty=\infty-\frac{1}{1-p}=\infty;$$

наконец, при $p=1$

$$I=\int_1^\infty \frac{1}{x}\,dx=\ln x\Big|_1^\infty=\ln\infty-\ln 1=\infty.$$

Итак, *интеграл* (51) *сходится при* $p>1$ *и расходится к бесконечности при* $p\leqslant 1$. Отсюда в силу признака (50) можно сделать то же заключение об интеграле (48), если

$$f(x)\sim\frac{A}{x^p}\quad \text{при}\quad x\to\infty.$$

Например, $\int_0^\infty \frac{dx}{\sqrt[3]{x^2+1}}=\infty$, так как $\frac{1}{\sqrt[3]{x^2+1}}=\frac{1}{x^{\frac{2}{3}}\sqrt{1+x^{-2}}}\sim\frac{1}{x^{\frac{2}{3}}}$; $p=\frac{2}{3}<1$;

$$\int_0^\infty \frac{dx}{\sqrt{x^3+1}}<\infty,\ \text{так как}\ \frac{1}{\sqrt{x^3+1}}\sim\frac{1}{x^{\frac{3}{2}}};\quad p=\frac{3}{2}>1; \tag{52}$$

$\int_0^\infty e^{-x^2}dx<\infty$, так как экспонента на бесконечности стремится к нулю быстрее любой степенной функции (п. IV.14) и можно применить признак сравнения (49).

Перейдем теперь к *интегралам от функции любого знака*

$$\int_a^\infty f(x)\,dx,\qquad f(x)\gtrless 0. \tag{53}$$

Здесь мы докажем только один признак: *если*

$$\int_a^\infty |f(x)|\,dx<\infty, \tag{54}$$

то интеграл (53) *сходится*; в этом случае он называется *абсолютно сходящимся*, а функция $f(x)$ называется *абсолютно интегрируемой* на интервале $a \leqslant x < \infty$.

Для доказательства следует ввести функции $f^+(x)$ и $f^-(x)$ — «положительную и отрицательную части» функции $f(x)$, которые изображены на рис. 266 и определяются следующим образом:

$$f^+(x) = \begin{cases} f(x) & \text{для тех } x, \text{ для которых } f(x) \geqslant 0, \\ 0 & \text{» » } x \text{ » » } f(x) < 0, \end{cases}$$

$$f^-(x) = \begin{cases} 0 & \text{» » } x \text{ » » } f(x) \geqslant 0, \\ |f(x)| & \text{» » } x \text{ » » } f(x) < 0. \end{cases}$$

Тогда можно написать

$$\begin{gathered} f(x) = f^+(x) - f^-(x), \\ |f(x)| = f^+(x) + f^-(x). \end{gathered} \tag{55}$$

Если условие (54) дано, то площадь, заштрихованная на рис. 266, *г*, конечная. Значит, конечны и площади, заштрихованные на

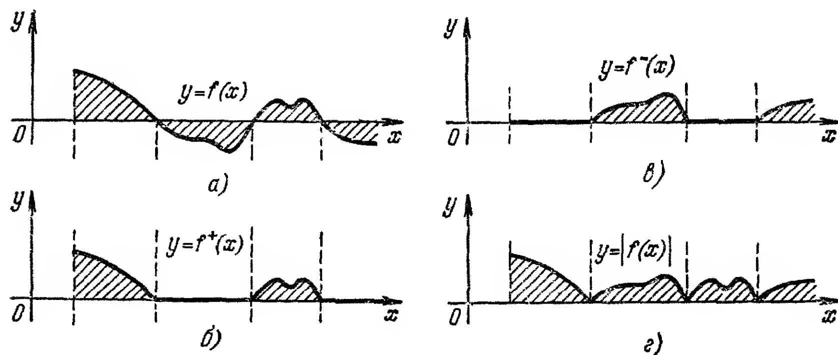


Рис. 266.

рис. 266, *б* и *в*, а так как $f^+(x) \geqslant 0$ и $f^-(x) \geqslant 0$, то несобственные интегралы от этих функций сходятся. Но теперь в силу равенства (55) и интеграл (53) сходится, что и требовалось доказать; более точно,

$$\int_a^\infty f(x)\,dx = \int_a^\infty f^+(x)\,dx - \int_a^\infty f^-(x)\,dx.$$

Может получиться, что интеграл (54) расходится, т. е. равен бесконечности, тогда как интеграл (53) сходится; это — *неабсолютная сходимость*. В этом случае площади, заштрихованные на рис. 266, *б* и *в*, бесконечны, но «сбалансированы» так, что если на рис. 266, *а*

учитывать знак площади, то бесконечности взаимно уничтожатся и останется конечный результат.

Например, интеграл

$$\int_1^\infty \frac{\sin x}{x^2}\,dx \tag{56}$$

сходится абсолютно, так как

$$\left|\frac{\sin x}{x^2}\right| \leqslant \frac{1}{x^2}.$$

и потому интеграл от абсолютной величины можно сравнить с интегралом (51) при $p=2$.

Если проделать то же с интегралом

$$\int_1^\infty \frac{\sin x}{x}\,dx, \tag{57}$$

то получится интеграл (51) при $p=1$, т. е. расходящийся. Тогда признак сравнения (см. неравенство (49)) непосредственно применить нельзя (почему?). Все же можно доказать, что

$$\int_1^\infty \left|\frac{\sin x}{x}\right| dx = \int_1^\infty |\sin x|\,\frac{1}{x}\,dx = \infty,$$

так как первый множитель периодически колеблется около *положительного* среднего значения. В то же время интеграл (57) сходится, так как после интегрирования по частям

$$\int_1^\infty \frac{\sin x}{x}\,dx = \left|\begin{matrix} \frac{1}{x}=u, & du=-\frac{1}{x^2}\,dx \\ \sin x\,dx=dv, & v=-\cos x \end{matrix}\right| =$$

$$= -\frac{\cos x}{x}\bigg|_1^\infty - \int_1^\infty \frac{\cos x}{x^2}\,dx = \cos 1 - \int_1^\infty \frac{\cos x}{x^2}\,dx,$$

а полученный интеграл того же типа, что и (56), и потому сходится. Итак, интеграл (57) сходится неабсолютно.

Все изложенное непосредственно переносится на интегралы от комплексных функций вещественного аргумента (п. VIII.6), а также на интегралы вида

$$\int_{-\infty}^b f(x)\,dx, \tag{58}$$

которые определяются с помощью соотношения

$$\int_{-\infty}^b f(x)\,dx = \lim_{-M\to-\infty} \int_{-M}^b f(x)\,dx$$

(впрочем, от интеграла (58) легко перейти к интегралу вида (41) с помощью подстановки $x=-y$).

16. Несобственные интегралы иных видов. Рассмотрим теперь несобственный интеграл вида

$$\int_a^b f(x)\,dx, \tag{59}$$

для которого пределы интегрирования конечны, но подынтегральная функция перестает быть конечной на одном из этих пределов, например, при $x \to a$. Чтобы придать такому интегралу смысл, опасный конец отрезают, после чего переходят к пределу:

$$\int_a^b f(x)\,dx = \lim_{\varepsilon \to +0} \int_{a+\varepsilon}^b f(x)\,dx.$$

Как и в п. 14, если конечного предела нет, то говорят, что интеграл (59) расходится.

Все свойства п. 15 непосредственно переносятся и на такие интегралы. Единственное отличие возникает при рассмотрении интегралов от степенных функций. Взамен интеграла (51) надо рассмотреть интеграл

$$\int_a^b \frac{1}{(x-a)^p}\,dx. \tag{60}$$

Легко проверить (проделайте это!), что *интеграл* (60) *сходится при* $p<1$ и *расходится при* $p \geqslant 1$.

Аналогично рассматривается интеграл (59), для которого подынтегральная функция перестает быть конечной на верхнем пределе интегрирования.

Точки на интервале интегрирования, в которых подынтегральная функция перестает быть конечной, а также бесконечные концы интервала интегрирования, называются *особенностями* рассматриваемого несобственного интеграла. До сих пор мы говорили об интегралах с единственной особенностью, лежащей в конце интервала интегрирования. Если особенность находится внутри интервала интегрирования или если особенностей несколько, то интегралу придают смысл по следующей схеме.

Пусть рассматривается интеграл

$$\int_a^b f(x)\,dx, \tag{61}$$

для которого подынтегральная функция перестает быть конечной в точках a, c и d, т. е. интеграл (61) имеет три особенности (на рис. 267 они изображены кружочками). Тогда при помощи точек

деления (на рис. 267 они изображены звездочками) разбивают интервал интегрирования на части так, чтобы на каждой была только одна особенность и притом в конце этой части; в рассматриваемом примере таких частей пять: $a\alpha$, αc, $c\beta$, βd, db. *Если каждый из интегралов*

$$\int_a^\alpha f(x)\,dx,\quad \int_\alpha^c f(x)\,dx,\quad \int_c^\beta f(x)\,dx,\quad \int_\beta^d f(x)\,dx,\quad \int_d^b f(x)\,dx \qquad (62)$$

сходится, то и интеграл (61) *считается сходящимся и равным сумме всех интегралов* (62). Если же хотя бы один из интегралов (62) расходится, то и интеграл (61) считается расходящимся и численного значения не имеет.

Рис. 267.

По описанной схеме определяется, в частности, интеграл вида $\int_{-\infty}^{\infty} f(x)\,dx$, где функция $f(x)$ конечна; здесь требуется одна точка деления.

К каждому из интегралов (62) можно применять признаки, описанные в п. 15 и 16. В частности, если

$$\int_a^b |f(x)|\,dx < \infty,$$

то интеграл (61) обязательно сходится; в этом случае он называется абсолютно сходящимся, а функция $f(x)$ — *абсолютно интегрируемой на интервале* a, b.

Остановимся отдельно на интеграле с двумя особенностями, лежащими в обоих концах интервала интегрирования. Если при этом к подынтегральной функции $f(x)$ удается найти первообразную $F(x)$, то вычисление интеграла можно производить следующим образом:

$$\int_a^b f(x)\,dx = \int_a^\alpha f(x)\,dx + \int_\alpha^b f(x)\,dx = \lim_{\varepsilon\to+0}\int_{a+\varepsilon}^\alpha f(x)\,dx + \lim_{\varepsilon\to+0}\int_\alpha^{b-\varepsilon} f(x)\,dx =$$

$$= \lim_{\varepsilon\to+0}[F(\alpha)-F(a+\varepsilon)] + \lim_{\varepsilon\to+0}[F(b-\varepsilon)-F(\alpha)] = F(b-0)-F(a+0).$$

Таким образом, в этом случае для вычисления определенного интеграла можно пользоваться обычной формулой (11), причем если подстановка пределов в первообразную даст конечный результат, то интеграл сходится.

Если интеграл (61) имеет особенность внутри интервала интегрирования, например в точке $x=c$, и известна первообразная $F(x)$, то

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx = [F(c-0)-F(a)] +$$

$$+ [F(b)-F(c+0)] = F(b)-F(a)+[F(c-0)-F(c+0)]. \qquad (63)$$

Значит, *если первообразная не имеет разрыва*, т. е. если $F(c-0)=F(c+0)$, *то при вычислении интеграла можно пользоваться обычной формулой* (11). Если первообразная имеет скачки, то надо делать поправку на эти скачки подобно (63). Если же первообразная имеет на интервале интегрирования разрывы более сложного вида, в частности, если она обращается в бесконечность, то интеграл расходится Эти правила справедливы при любом числе особенностей.

Например, интеграл $\int\limits_{-1}^{2}\frac{1}{\sqrt[3]{x}}dx$ можно вычислить так:

$$\int\limits_{-1}^{2}\frac{1}{\sqrt[3]{x}}dx=\int\limits_{-1}^{2}x^{-\frac{1}{3}}dx=\left.\frac{x^{\frac{2}{3}}}{\frac{2}{3}}\right|_{-1}^{2}=\frac{3}{2}(2^{\frac{2}{3}}-1)=0{,}881,$$

так как в данном случае первообразная, пропорциональная $x^{\frac{2}{3}}$, непрерывна и потому интеграл сходится. В то же время последний пример в п. 4 был рассчитан неверно, так как там первообразная, т. е. $-\frac{1}{x}$, обращалась в бесконечность на интервале интегрирования при $x=0$ и потому интеграл расходился.

При вычислении несобственных интегралов широко применяются разложения в ряды различного вида Если такое разложение хорошо действует только вблизи особенности, то заданный интеграл представляют в виде суммы собственного и несобственного, взятого по интервалу около особенности; первый вычисляют по методам § 3, а второй разлагают в ряд. Например, для подсчета интеграла (52) можно поступить так:

$$\int\limits_{0}^{\infty}\frac{dx}{\sqrt{x^3+1}}=\int\limits_{0}^{a}\frac{dx}{\sqrt{x^3+1}}+\int\limits_{a}^{\infty}x^{-\frac{3}{2}}\left(1+\frac{1}{x^3}\right)^{-\frac{1}{2}}dx=$$

$$=\int\limits_{0}^{a}\frac{dx}{\sqrt{x^3+1}}+\int\limits_{a}^{\infty}\left(x^{-\frac{3}{2}}-\frac{1}{2}x^{-\frac{9}{2}}+\frac{3}{8}x^{-\frac{15}{2}}-\ldots\right)dx=$$

$$=\int\limits_{0}^{a}\frac{dx}{\sqrt{x^3+1}}+\frac{2}{\sqrt{a}}-\frac{1}{7\sqrt{a^7}}+\frac{3}{52\sqrt{a^{13}}}-\ldots=\int\limits_{0}^{a}\frac{dx}{\sqrt{x^3+1}}+S; \qquad (64)$$

при этом для разложения подынтегральной функции мы воспользовались формулой бинома Ньютона (IV 60) Здесь $a>0$ — произвольное число; но если a взять слишком большим, то будет трудно подсчитывать последний интеграл, а если слишком маленьким, то будет плохо с рядом. Возьмем $a=2$ и подсчитаем последний интеграл по формуле Симпсона (п 13), разбив интервал интегрирования на восемь частей. Тогда получим (проверьте!)

$$\int\limits_{0}^{2}\frac{dx}{\sqrt{x^3+1}}=1{,}402; \qquad S=1{,}402,$$

и тем самым интеграл (52) равен 2,804.

Для расходящегося интеграла, например, вида (41) может возникнуть вопрос о более точной характеристике поведения его «конечной части» (42) при «исчерпывании особенности», т. е. при $N\longrightarrow\infty$. Для выяснения этого

также применяются разложения в ряды. Например,

$$\int_0^N \frac{dx}{\sqrt[3]{x^2+1}} = \int_0^a \frac{dx}{\sqrt[3]{x^2+1}} + \int_a^N x^{-\frac{2}{3}}(1+x^{-2})^{-\frac{1}{3}}\,dx =$$

$$= \int_0^a \frac{dx}{\sqrt[3]{x^2+1}} + \int_a^N \left(x^{-\frac{2}{3}} - \frac{1}{3}x^{-\frac{8}{3}} + \frac{2}{9}x^{-\frac{14}{3}} - \ldots\right)dx =$$

$$= 3N^{\frac{1}{3}} + C + \frac{1}{5}N^{-\frac{5}{3}} - \frac{2}{33}N^{-\frac{11}{3}} + \ldots,$$

где $C = \int_0^a \frac{dx}{\sqrt[3]{x^2+1}} - 3a^{\frac{1}{3}} - \frac{1}{5}a^{-\frac{5}{3}} + \frac{2}{33}a^{-\frac{11}{3}} - \ldots$ — постоянная, которую можно вычислить, как в предыдущем примере.

Аналогичный вопрос может возникнуть для сходящегося интеграла. Так, рассуждая, как при вычислениях (64), получим

$$\int_0^N \frac{dx}{\sqrt{x^3+1}} = \int_0^\infty \frac{dx}{\sqrt{x^3+1}} - \int_N^\infty \frac{dx}{\sqrt{x^3+1}} = 2,801 - \frac{2}{\sqrt{N}} + \frac{1}{7\sqrt{N^7}} - \ldots;$$

интегрируя по частям, найдем

$$\int_0^N e^{-x^2}dx = \int_0^\infty e^{-x^2}dx - \int_N^\infty \frac{1}{2x}\cdot 2xe^{-x^2}dx =$$

$$= \int_0^\infty e^{-x^2}dx + \frac{1}{2x}e^{-x^2}\Big|_N^\infty + \frac{1}{2}\int_N^\infty \frac{e^{-x^2}}{x^2}\,dx =$$

$$= \int_0^\infty e^{-x^2}dx - \frac{1}{2N}e^{-N^2} + \text{величина порядка } \frac{1}{N^3}e^{-N^2}.$$

То, что полученный интеграл имеет указанный порядок, легко доказать с помощью правила Лопиталя, что мы предоставляем читателю, для уточнения разложения можно применить повторное интегрирование по частям.

17. Гамма-функция. В качестве важного примера несобственного интеграла рассмотрим неэлементарную *«гáмма-функцию»*, введенную Эйлером в 1729 г.:

$$\Gamma(p) = \int_0^\infty e^{-x}x^{p-1}dx. \tag{65}$$

Этот интеграл называется также *эйлеровым интегралом второго рода*.

Интеграл (65) несобственный уже из-за бесконечного верхнего предела. Однако так как e^{-x} при $x \to \infty$ стремится к нулю быстрее любой степени x, то на верхнем пределе интеграл (65) сходится. Если $p < 1$, то интеграл (65) имеет особенность и при $x = 0$. Так как $e^{-x}x^{p-1} \sim \frac{1}{x^{1-p}}$ при $x \to 0$, то в силу начала п. 16 интеграл (65) сходится при $1 - p < 1$, т. е. при $p > 0$, и

расходится при $1-p\geqslant 1$, т. е. при $p\leqslant 0$. Итак, *формулу* (65) *надо рассматривать при* $0<p<\infty$.

Для вывода основного свойства гамма-функции произведем интегрирование по частям:

$$\Gamma(p+1)=\int_0^\infty e^{-x}x^{(p+1)-1}\,dx=\int_0^\infty e^{-x}x^p dx=-e^{-x}x^p\Big|_0^\infty+\int_0^\infty e^{-x}px^{p-1}dx.$$

Выделившийся член равен нулю на обоих пределах (почему?), и мы получаем важную формулу:

$$\Gamma(p+1)=p\Gamma(p). \tag{66}$$

Легко подсчитать, далее,

$$\Gamma(1)=\int_0^\infty e^{-x}x^{1-1}\,dx=\int_0^\infty e^{-x}dx=-e^{-x}\Big|_0^\infty=1.$$

Если теперь в формулу (66) подставлять последовательно $p=1, 2, 3, \ldots$, мы получим

$\Gamma(2)=1\cdot\Gamma(1)=1$; $\Gamma(3)=2\Gamma(2)=2\cdot 1$; $\Gamma(4)=3\Gamma(3)=3\cdot 2\cdot 1$ и т. д.; вообще

$$\Gamma(n+1)=n! \qquad (n=1,\ 2,\ 3,\ \ldots). \tag{67}$$

Если прочитать формулу (67) справа налево, то мы видим, что гамма-функция дает представление факториала. В то же время она имеет смысл и для нецелых значений аргумента и тем самым продолжает факториальную функцию (п. I.15) с дискретных значений аргумента на непрерывные. График этой функции изображен на рис. 268; на нем, в частности, показано равенство $\Gamma(+0)=+\infty$, вытекающее из формулы (66).

Рис. 268.

Из формулы (67) видно, в частности, что

$$0!=\Gamma(1)=1.$$

Далее,

$$\left(-\frac{1}{2}\right)!=\Gamma\left(-\frac{1}{2}+1\right)=\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}=1{,}772;$$

это значение будет выведено в п. 18 (формула (71)). Отсюда

$$\left(\frac{1}{2}\right)!=\Gamma\left(\frac{3}{2}\right)=\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{\sqrt{\pi}}{2}=0{,}886;\ \left(\frac{3}{2}\right)!=\Gamma\left(\frac{5}{2}\right)=\frac{3}{2}\Gamma\left(\frac{3}{2}\right)=\frac{3\sqrt{\pi}}{4}=1{,}329 \text{ и т. д.}$$

Гамма-функция определяется и *при отрицательных значениях аргумента*. Пользоваться формулой (65) при этом нельзя, так как интеграл расходится. Однако можно применить формулу (66), переписав ее в виде

$$\Gamma(p)=\frac{\Gamma(p+1)}{p}. \tag{68}$$

Если $-1 < p < 0$, то $0 < p+1 < 1$, и потому правая часть имеет смысл, чем и определяется значение $\Gamma(p)$ при таких p; отметим, что получится $\Gamma(p) < 0$. Если, далее, $-2 < p < -1$, то $-1 < p+1 < 0$, и потому правая часть равенства (68) уже определена, чем определяется и левая, причем $\Gamma(p) > 0$. Затем определяем $\Gamma(p)$ при $-3 < p < -2$ и т. д. Тем самым $\Gamma(p)$ определится для p любого знака, причем для всех p имеет место формула (66). Из формулы (68) видно также последовательно, что $\Gamma(0) = \pm\infty$, $\Gamma(-1) = \pm\infty$, $\Gamma(-2) = \pm\infty$ и т. д. График гамма-функции при отрицательных значениях аргумента показан на рис. 269.

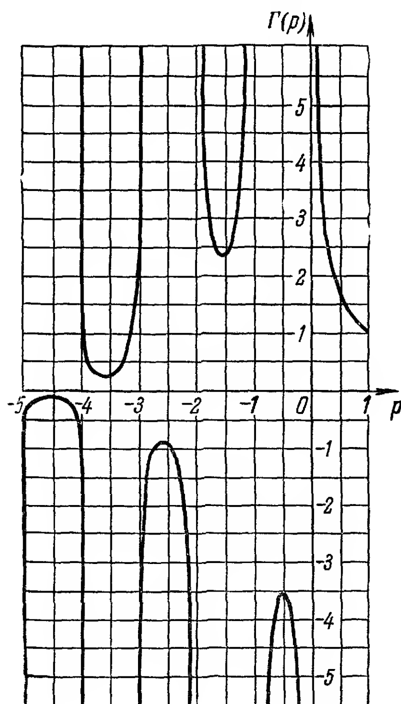


Рис. 269.

Таблицы значений гамма-функции имеются, в частности, в книге [53].

18. Бета-функция. *Бе́та-функция, или эйлеров интеграл первого рода*, определяется формулой

$$B(p,\ q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx. \tag{69}$$

При этом должно быть $p > 0$ и $q > 0$, так как в противном случае интеграл расходится на нижнем или на верхнем пределе (почему?). Отметим, что неопределенный интеграл (69) берется элементарно только для специальных комбинаций показателей p и q (п. XIII.9).

Как будет выведено в п. XVI.17, *бета-функция выражается через гамма-функцию по формуле*

$$B(p,\ q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}. \tag{70}$$

Из этой формулы вытекает интересное следствие: положив $p = q = \frac{1}{2}$, получим

$$\Gamma^2\left(\frac{1}{2}\right) = B\left(\frac{1}{2},\ \frac{1}{2}\right) = \int_0^1 x^{-\frac{1}{2}}(1-x)^{-\frac{1}{2}}\,dx =$$

$$= \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = 2\int_0^1 \frac{dx}{\sqrt{1-(1-2x)^2}} = -\arcsin(1-2x)\Big|_0^1 = \pi,$$

а так как $\Gamma(p) > 0$ при $p > 0$, то

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{71}$$

Отсюда в свою очередь можно вывести значение важного интеграла

$$\int_0^\infty e^{-x^2}\,dx = \left| x=\sqrt{t} \right| = \frac{1}{2}\int_0^\infty e^{-t} t^{-\frac{1}{2}}\,dt =$$

$$= \frac{1}{2}\int_0^\infty e^{-t} t^{\frac{1}{2}-1}\,dt = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}. \qquad (72)$$

Через бета- и тем самым через гамма-функцию выражаются многие определенные интегралы, для которых неопределенные интегралы при произвольных показателях не берутся в элементарных функциях, например,

$$\int_0^{\frac{\pi}{2}} \sin^p x\,dx = \left| \sin x = \sqrt{t} \right| = \int_0^1 \frac{1}{2} t^{\frac{p}{2}-\frac{1}{2}} (1-t)^{-\frac{1}{2}}\,dt =$$

$$= \frac{1}{2}\mathrm{B}\left(\frac{p+1}{2}, \frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}\,\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}+1\right)} \quad (p>-1);$$

$$\int_0^\infty \frac{x^{p-1}}{(1+x)^{p+q}}\,dx = \left| x = \frac{y}{1-y} \right| = \int_0^1 \frac{y^{p-1}(1-y)^{p+q}}{(1-y)^{p-1}}\,\frac{dy}{(1-y)^2} =$$

$$= \int_0^1 y^{p-1}(1-y)^{q-1}\,dy = \mathrm{B}(p, q) \qquad (p>0,\ q>0); \qquad (73)$$

$$\int_0^\infty \frac{dx}{(1+x^p)^q} = \left| x = y^{\frac{1}{p}} \right| = \frac{1}{p}\int_0^\infty \frac{y^{\frac{1}{p}-1}}{(1+y)^q}\,dy =$$

$$= \frac{1}{p}\mathrm{B}\left(\frac{1}{p}, q-\frac{1}{p}\right) \qquad (p>0,\ qp>1) \qquad (74)$$

(см. (73)) и т. д. В частности, из (74) получаем значение интеграла (52):

$$\int_0^\infty \frac{dx}{\sqrt{1+x^3}} = \frac{1}{3}\mathrm{B}\left(\frac{1}{3}, \frac{1}{2}-\frac{1}{3}\right) = \frac{1}{3}\mathrm{B}\left(\frac{1}{3}, \frac{1}{6}\right) =$$

$$= \frac{1}{3}\,\frac{\Gamma\left(\frac{1}{3}\right)\cdot\Gamma\left(\frac{1}{6}\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{2{,}676\cdot 5{,}566}{3\sqrt{\pi}} = 2{,}801.$$

19. Главное значение расходящегося интеграла. В некоторых случаях, в частности при исследовании сплошных сред, оказывается целесообразным отдельным расходящимся интегралам приписывать в некотором условном смысле числовое значение. Это можно сделать различными способами. Коши

предложил делать это так. Пусть интеграл

$$\int_a^b f(x)\,dx \tag{75}$$

имеет единственную особенность *внутри* интервала интегрирования в точке $x=c$. Тогда эту особенность *симметрично* вырезают, после чего переходят к пределу, т. е. полагают

$$\text{v. p.}\int_a^b f(x)\,dx = \lim_{\varepsilon\to+0}\left[\int_a^{c-\varepsilon} f(x)\,dx + \int_{c+\varepsilon}^b f(x)\,dx\right]. \tag{76}$$

Такой предел может существовать и в том случае, если интеграл (75) в обычном смысле (п. 16) расходится. Тогда предел (76) называется *главным значением* интеграла (75), которое обозначается буквами v. p. (от английского value principal, главное значение) при знаке интеграла, а сам интеграл часто называют *сингулярным*, в отличие от собственных или сходящихся несобственных интегралов, которые (те и другие) называют *регулярными*. Подобным образом главным значением интеграла, взятого по всей числовой оси, называется предел

$$\text{v. p.}\int_{-\infty}^{\infty} f(x)\,dx = \lim_{N\to\infty}\int_{-N}^{N} f(x)\,dx.$$

Например, интеграл $\int_{-1}^{2}\frac{1}{x}\,dx$ расходится, так как первообразная функция, т. е. $\ln|x|$, имеет бесконечный разрыв на интервале интегрирования при $x=0$. В то же время главное значение

$$\text{v. p.}\int_{-1}^{2}\frac{1}{x}\,dx = \lim_{\varepsilon\to+0}\left[\int_{-1}^{-\varepsilon}\frac{1}{x}\,dx + \int_{\varepsilon}^{2}\frac{1}{x}\,dx\right] =$$

$$= \lim_{\varepsilon\to+0}\left[\ln|x|\,\Big|_{-1}^{-\varepsilon} + \ln|x|\,\Big|_{\varepsilon}^{2}\right] = \lim_{\varepsilon\to+0}[\ln\varepsilon - \ln 1 + \ln 2 - \ln\varepsilon] = \ln 2 = 0{,}693$$

существует, так как опасные слагаемые $\pm\ln\varepsilon$ взаимно уничтожаются до перехода к пределу. Другой пример:

$$\text{v. p.}\int_{-\infty}^{\infty}\sin x\,dx = \lim_{N\to\infty}\int_{-N}^{N}\sin x\,dx = \lim_{N\to\infty}(-\cos x)\Big|_{-N}^{N} =$$

$$= \lim_{N\to\infty}[-\cos N + \cos N] = 0,$$

т. е. и здесь главное значение интеграла существует, хотя сам интеграл расходится и потому является сингулярным.

Конечно, далеко не все расходящиеся интегралы обладают главным значением.

§ 5. Интегралы, зависящие от параметра

20. Собственные интегралы. Рассмотрим интеграл вида

$$I=\int_a^b f(x;\ \lambda)\,dx, \tag{77}$$

где под знак интеграла, помимо переменной интегрирования x, входит *параметр* (произвольная постоянная) λ, т. е. величина, которая в процессе интегрирования считается постоянной, но вообще может принимать различные значения. Тогда и результат интегрирования, вообще говоря, зависит от λ, т. е. $I=I(\lambda)$. Такие интегралы часто встречаются в приложениях, когда интегрируемая функция включает в себя какие-либо массы, размеры и т. п., которые в процессе интегрирования являются постоянными. Мы для простоты будем считать, что подынтегральная функция содержит только один параметр, хотя результаты получаются аналогичными при любом числе параметров.

Приведем несколько формальных примеров:

$$\int_0^1 (x^2+\lambda x)\,dx=\frac{1}{3}+\frac{\lambda}{2};\quad \int_0^\pi \sin\alpha x\,dx=\frac{1-\cos\alpha}{\alpha};$$

$$\int_0^1 (s+1)\,x^s dx=1 \qquad (s>-1).$$

Мы будем считать здесь интеграл (77) собственным, т. е. пределы интегрирования и подынтегральную функцию конечными, и рассмотрим некоторые его свойства.

1. *Если подынтегральная функция при $a\leqslant x\leqslant b$ зависит от λ непрерывно, то и интеграл I зависит от λ непрерывно.* Это вытекает, например, из геометрического смысла интеграла как площади криволинейной трапеции: если при бесконечно малом изменении λ криволинейная сторона трапеции изменится бесконечно мало, то и площадь изменится бесконечно мало.

Отметим, что при этом функция f не обязана зависеть от x непрерывно; она может иметь конечные разрывы.

Бывает, что и пределы интегрирования зависят от параметра:

$$I(\lambda)=\int_{a(\lambda)}^{b(\lambda)} f(x;\ \lambda)\,dx. \tag{78}$$

Тогда для непрерывности $I(\lambda)$ надо дополнительно потребовать, чтобы функции $a(\lambda)$ и $b(\lambda)$ не имели разрывов.

2. *Правило Лейбница: возможно дифференцирование по параметру под знаком интеграла* (77), другими словами,

$$\frac{dI}{d\lambda}=\int_a^b f'_\lambda(x;\ \lambda)\,dx. \tag{79}$$

Дело в том, что интеграл (77) аналогичен сумме весьма большого числа слагаемых (п. 2), каждое из которых зависит от λ, а дифференцирование под знаком суммы возможно, так как производная суммы равна сумме производных.

Формула (79) понимается в простейшем смысле, если не только интеграл (77) собственный, но и интеграл (79) собственный, или несобственный, но сходящийся. Бывает, что интеграл (79) расходится: тогда формула (79) все же справедлива, но в некотором обобщенном смысле, о котором будет сказано в п. 27.

При дифференцировании интеграла (78) надо учесть, что λ в правую часть входит трижды, так что надо пользоваться формулой для производной сложной функции (п. IX.12), а также формулами для производной от интеграла по верхнему и нижнему пределам (п. 4). Получится

$$\frac{dI}{d\lambda}=\int\limits_{a(\lambda)}^{b(\lambda)} f'_\lambda(x;\ \lambda)\,dx+f(b(\lambda);\ \lambda)b'(\lambda)-f(a(\lambda);\ \lambda)\,a'(\lambda). \qquad (80)$$

3. *Возможно интегрирование по параметру под знаком интеграла* (77):

$$\int\limits_\alpha^\beta I(\lambda)\,d\lambda=\int\limits_a^b\left(\int\limits_\alpha^\beta f(x;\ \lambda)\,d\lambda\right)dx.$$

Это обосновывается так же, как свойство 2.

21. Несобственные интегралы. Мы рассмотрим интеграл вида

$$I(\lambda)=\int\limits_a^\infty f(x;\ \lambda)\,dx, \qquad (81)$$

не имеющий особенностей при конечных x; свойства несобственных интегралов иных видов (п. 16) аналогичны. Конечно, прежде всего надо требовать, чтобы интеграл (81) сходился. Однако по сравнению с п. 20 мы сталкиваемся со следующим новым обстоятельством: *даже если функция f непрерывно зависит от λ, зависимость интеграла от λ может получиться разрывной*. Это связано с тем, что бесконечно малое изменение функции на бесконечно большом участке интегрирования может привести к конечному изменению интеграла.

Например, в п. XVII.32 мы покажем, что

$$\int\limits_0^\infty \frac{\sin x}{x}\,dx=\frac{\pi}{2}.$$

Отсюда сразу следует, что при $\lambda>0$

$$I(\lambda)=\int\limits_0^\infty \frac{\sin\lambda x}{x}\,dx=|\,\lambda x=s\,|=\int\limits_0^\infty \frac{\sin s}{s}\,ds=\frac{\pi}{2}.$$

В то же время при $\lambda=0$ получается $I=0$, а при $\lambda<0$

$$I(\lambda)=\int\limits_0^\infty \frac{\sin(-|\lambda|)\,x}{x}\,dx=-\int\limits_0^\infty \frac{\sin|\lambda|\,x}{x}\,dx=-\frac{\pi}{2}.$$

Итак, в данном примере $I(\lambda)$ при $\lambda=0$ имеет скачок. Это может показаться странным, так как несобственный интеграл получается как предел собственных,

$$I(\lambda)=\lim_{N\to\infty}\int\limits_0^N \frac{\sin\lambda x}{x}\,dx,$$

а каждый собственный интеграл зависит от λ непрерывно. Но дело в том, что *предел непрерывных функций*, как мы сейчас увидим, *не обязан быть непрерывной функцией*!

На рис. 270 показаны графики функций

$$I_N(\lambda)=\int_0^N \frac{\sin\lambda x}{x}\,dx$$

при малом и при большом N. Хотя эти функции и непрерывные, но при большом N переход от $-\frac{\pi}{2}$ к $+\frac{\pi}{2}$ совершается на малом интервале λ, причем чем больше N, тем этот интервал меньше. В пределе, при $N=\infty$, этот переход осуществляется на бесконечно малом интервале λ, т. е. появляется разрыв.

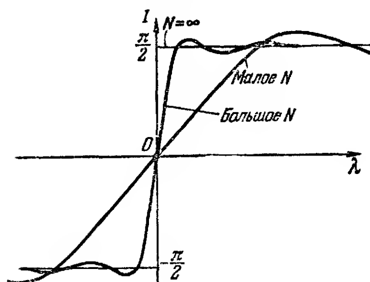


Рис. 270.

Возможность таких разрывов усложняет исследование интегралов вида (81), в частности, применение свойств 2 и 3 из п. 20. Поэтому иногда взамен (81) приходится рассматривать интеграл по конечному интервалу, от a до N, а затем переходить к пределу при $N\to\infty$. Тем не менее как мы сейчас покажем, имеется важный частный случай, когда такие разрывы невозможны.

Назовем интеграл (81) *правильно сходящимся*, если при всех рассматриваемых значениях λ

$$|f(x;\ \lambda)|\leqslant\varphi(x)\quad(a\leqslant x<\infty),\ \text{ где } \int_a^\infty \varphi(x)\,dx<\infty. \tag{82}$$

Например, интеграл $\int_1^\infty \frac{\sin\lambda x}{x^2}\,dx$ сходится правильно, так как

$$\left|\frac{\sin\lambda x}{x^2}\right|\leqslant\frac{1}{x^2},\qquad \int_1^\infty \frac{1}{x^2}\,dx=1<\infty.$$

Относительно *правильно сходящегося интеграла* (81) можно утверждать, что *если подынтегральная функция зависит от λ непрерывно, то и интеграл I зависит от λ непрерывно*. Действительно, этот интеграл можно представить в виде

$$I(\lambda)=\int_a^N f(x;\ \lambda)\,dx+\int_N^\infty f(x;\ \lambda)\,dx.$$

Здесь первое слагаемое, как собственный интеграл, зависит от λ непрерывно, тогда как второе можно оценить так:

$$\left|\int_N^\infty f(x;\ \lambda)\,dx\right|\leqslant\int_N^\infty |f(x;\ \lambda)|\,dx\leqslant\int_N^\infty \varphi(x)\,dx,$$

и в силу условия (82) этот интеграл при достаточно большом N будет весьма мал *сразу для всех значений* λ (п. 15). Поэтому малому изменению λ отвечает малое изменение и всей суммы $I(\lambda)$, что и означает непрерывную зависимость I от λ.

Свойства правильно сходящихся интегралов полностью аналогичны свойствам собственных интегралов, описанным в п. 20.

§ 6. Криволинейные интегралы

22. Интеграл по длине дуги. Третий пример п. 1 является одновременно простейшим примером криволинейного интеграла по длине дуги. В общем случае определение дается следующим образом.

Пусть в пространстве или на плоскости дана конечная линия (L) и в каждой ее точке задано значение некоторой величины u. Если отсчитывать вдоль (L) дугу s от некоторой точки, то можно считать, что u является функцией s, $u=f(s)$. Чтобы составить интегральную сумму, надо мысленно разбить линию (L) на маленькие примыкающие друг к другу дуги; эта сумма имеет вид

$$\sum_{k=1}^{n} f(\sigma_k)\,\Delta s_k \qquad (\alpha=s_0<s_1<\ldots<s_n=\beta, \quad s_{k-1}\leqslant\sigma_k\leqslant s_k),$$

где значения $s=\alpha$ и $s=\beta$ отвечают концам линии (L), а n — число участков разбиения. Чтобы получить интеграл, надо перейти к пределу в процессе, когда разбиение бесконечно измельчается (ср. п. 2, но без геометрического смысла):

$$\int_{(L)} u\,ds=\int_{(L)} f(s)\,ds=\int_{\alpha}^{\beta} f(s)\,ds=\lim\sum_{k=1}^{n} f(\sigma_k)\,\Delta s_k. \tag{83}$$

Этот интеграл и называется *криволинейным интегралом по длине дуги*. Таким образом, в упомянутом примере п. 1

$$M=\int_{(L)} \rho\,ds.$$

Подобным образом (ср. пример п. 6), если точка проходит линию (L), причем на нее действует сила $\mathbf{F}$, вообще говоря переменная, то работа

$$A=\int_{(L)} F\cos(\widehat{\mathbf{F},\ \boldsymbol{\tau}})\,ds,$$

где $\boldsymbol{\tau}$ — единичный вектор («орт») касательной к (L).

Так как из формулы (83) видно, что криволинейный интеграл по длине дуги — это разновидность обычного определенного интеграла, то многие свойства определенного интеграла (см., в частности, свойства 2—5 п. 4, а также п. 5) автоматически распространяются на криволинейный интеграл. В то же время надо иметь в виду, что Δs и потому ds здесь *считаются всегда положительными*, т. е.

при переходе в (83) к определенному интегралу интегрирование всегда идет от меньшего к большему; поэтому свойство 1 п. 4 для криволинейных интегралов по дуге лишено смысла.

Иногда величина u задается во всем пространстве, например, $u=f(x, y, z)$. Тогда интеграл (83) можно записать в виде

$$I=\int_{(L)} f(x,y,z)\,ds.$$

Если линия (L) задана в параметрическом виде $x=x(t)$, $y=y(t)$, $z=z(t)$, то вычисление этого интеграла можно производить по формуле

$$I=\int_{\gamma}^{\delta} f(x(t),y(t),z(t))\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\,dt,$$

где значения $t=\gamma$ и $t=\delta$ отвечают концам линии (L); выражение для ds мы взяли из п. VII.23. На основании того же пункта интеграл по дуге иногда записывают в виде $\int_{(L)} u\,|d\mathbf{r}|$. Пишут и так: $\int_{(L)} u(M)\,ds$, где M — текущая точка линии (L); тогда и интегральную сумму можно записать в виде

$$\sum_{k=1}^{n} u_k\Delta s_k=\sum_{k=1}^{n} u(M_k)\,\Delta s_k,$$

где M_k — некоторая точка на k-й дуге разбиения.

Покажем пример применения интеграла по длине дуги. Из механики известно, что если в плоскости дана система материальных точек $M_k(x_k;\ y_k)$ массы m_k, где $k=1, \ldots, n$, то координаты центра тяжести этой системы определяются по формулам

$$x_{\text{ц. т.}}=\frac{m_1x_1+m_2x_2+\ldots+m_nx_n}{m_1+m_2+\ldots+m_n}, \quad y_{\text{ц. т.}}=\frac{m_1y_1+m_2y_2+\ldots+m_ny_n}{m_1+m_2+\ldots+m_n}.$$

Если теперь представить себе в плоскости материальную линию (L) с, вообще говоря, непостоянной линейной плотностью ρ, то центр тяжести этой линии можно найти следующим образом.

Разобьем мысленно (L) на маленькие дуги Δs_k и заменим каждую из дуг материальной точкой массы $m_k=\rho_k\Delta s_k$, расположенной на этой дуге. Эта «дискретная модель» материальной линии имеет центр тяжести с координатами

$$x_{\text{ц. т.}}=\frac{\sum\limits_{k=1}^{n} x_k\rho_k\Delta s_k}{\sum\limits_{k=1}^{n} \rho_k\Delta s_k}, \quad y_{\text{ц.т.}}=\frac{\sum\limits_{k=1}^{n} y_k\rho_k\Delta s_k}{\sum\limits_{k=1}^{n} \rho_k\Delta s_k}. \tag{84}$$

Если перейти к пределу при бесконечном измельчении разбиения, то дискретная модель перейдет в непрерывную линию, для которой в силу формул (84) получим

$$x_{\text{ц. т.}}=\frac{\int\limits_{(L)}\rho x\,ds}{\int\limits_{(L)}\rho\,ds},\qquad y_{\text{ц. т.}}=\frac{\int\limits_{(L)}\rho y\,ds}{\int\limits_{(L)}\rho\,ds}. \tag{85}$$

Особенно простые формулы получатся для линии с постоянной плотностью; тогда центр тяжести называется *геометрическим*. Из формул (85), сокращая на ρ, выводим

$$x_{\text{г. ц. т.}}=\frac{\int\limits_{(L)}x\,ds}{L},\quad y_{\text{г. ц. т.}}=\frac{\int\limits_{(L)}y\,ds}{L}, \tag{86}$$

где под L понимается длина линии (L).

Если сравнить вторую формулу (86) с формулой (36) для площади поверхности вращения, то мы увидим, что

$$S=2\pi\int\limits_{(L)}y\,ds=L\cdot 2\pi y_{\text{г. ц. т.}}$$

Другими словами, *если плоская линия вращается вокруг оси, лежащей в плоскости этой линии и не пересекающей ее, то площадь полученной поверхности вращения равна произведению длины этой линии на путь, пройденный ее геометрическим центром тяжести.* Эта теорема называется «*первой теоремой Гюльдена*» по имени применявшего ее швейцарского математика П. Гюльдéна (1577—1643). Ее применение особенно удобно, если положение центра тяжести легко определить. Например, из нее получаем площадь поверхности тора (рис. 271), полученной вращением окружности вокруг оси, лежащей в плоскости этой окружности и не пересекающей ее,

$$S=2\pi r\cdot 2\pi R=4\pi^2 rR.$$

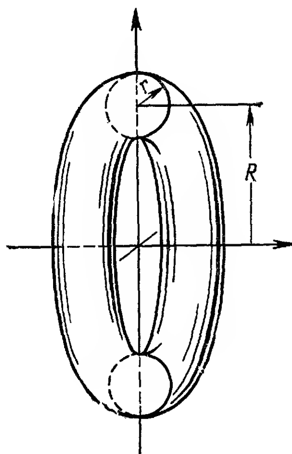


Рис. 271.

К понятию криволинейного интеграла по длине дуги мы еще вернемся в п. XVI.1.

23. Интеграл по координате. Помимо криволинейных интегралов по длине дуги, рассматриваются также *криволинейные интегралы по координате*. При составлении последнего линия (L) предполагается ориентированной, т. е. должно быть указано, в каком направлении она проходится; если линия разомкнутая, это значит, что надо указать, какая из граничных точек этой линии считается ее началом, а какая—концом. Для определения интеграла надо взамен

формулы (83) написать

$$\left.\begin{aligned} &\int\limits_{(L)} u\,dx = \lim \sum_{k=1}^{n} f(\sigma_k)\,\Delta x_k,\\ &\int\limits_{(L)} u\,dy = \lim \sum_{k=1}^{n} f(\sigma_k)\,\Delta y_k, \quad \int\limits_{(L)} u\,dz = \lim \sum_{k=1}^{n} f(\sigma_k)\,\Delta z_k, \end{aligned}\right\} \qquad (87)$$

где Δx_k — приращение абсциссы x на k-й дуге разбиения и т. д.

Интегралы вида (87) легко сводятся к обычным определенным интегралам. Так, если линия (L) задана в параметрическом виде, то и значения u вдоль (L) становятся функцией параметра t и

$$\int\limits_{(L)} u\,dx = \int\limits_{\gamma}^{\delta} u(t)\,\dot{x}(t)\,dt,$$

где значения $t=\gamma$ и $t=\delta$ отвечают концам линии (L). Поэтому основные свойства определенного интеграла (свойства 2—5 п. 4) распространяются и на интегралы по координате. В то же время здесь имеет место и свойство 1 п. 4, которое можно сформулировать так: *при перемене ориентации линии* (L) *интегралы* (87) *умножаются на* -1; действительно, если линию (L) проходить в противоположном направлении, то все Δx, а с ними и dx, поменяют знак. Возможность перемены знака у dx приводит также к тому, что на интегралы по координате не распространяются свойства, связанные с интегрированием неравенств (п. 5): так, интеграл вида (87) от положительной функции, в отличие от интеграла по длине дуги, не обязан быть положительным.

В теории дифференциальных уравнений (п. XV.6) и в теории векторного поля (§ XVI.6) применяются *комбинации интегралов* (86) вида

$$\int\limits_{(L)} (u\,dx + v\,dy + w\,dz) = \int\limits_{(L)} u\,dx + \int\limits_{(L)} v\,dy + \int\limits_{(L)} w\,dz.$$

Примеры криволинейных интегралов по координатам, по существу, были в п. 8, так как формулы (29)—(31) в силу равенств $\dot{x}\,dt = dx$, $\dot{y}\,dt = dy$ можно переписать так:

$$S = -\int\limits_{(L)} y\,dx = \int\limits_{(L)} x\,dy = \frac{1}{2}\int\limits_{(L)} (x\,dy - y\,dx). \qquad (88)$$

В дальнейшем нам понадобится также интеграл $\int\limits_{(L)} y\,dx$, распространенный по плоской замкнутой линии (L), произвольно расположенной в пространстве (рис. 272). Для его вычисления спроектируем линию (L) на плоскость xOy и заметим, что

$$\int\limits_{(L)} y\,dx = \int\limits_{(L')} y\,dx, \qquad (89)$$

так как соответствующие точки линий (L) и (L') отличаются только значениями z, которые никак не проявляются в интегралах (89).

Для правого интеграла (89) в силу (88)

$$\int_{(L')} y\,dx = -S'.$$

Воспользуемся теперь тем, что *при проектировании плоской фигуры на другую плоскость площадь проекции равна произведению площади исходной*

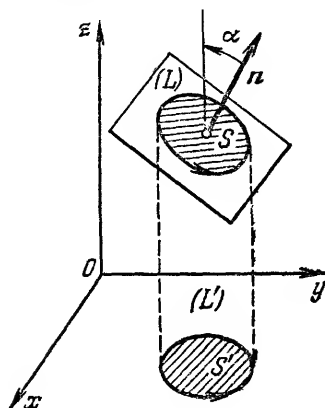


Рис. 272.

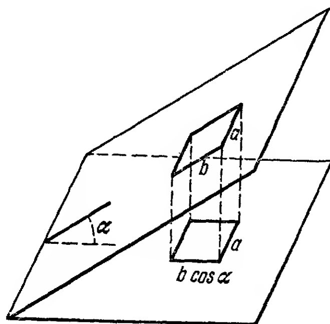


Рис. 273.

фигуры на косинус угла между обеими плоскостями, так как при этом (рис 273) размеры в одном направлении не меняются, тогда как в другом множатся на $\cos\alpha$. Итак,

$$\int_{(L)} y\,dx = -S\cos\alpha = -S\cos(\widehat{\mathbf{n}, z}), \tag{90}$$

где S — площадь, ограниченная линией (L), а $\mathbf{n}$ — единичный вектор внешней нормали к плоскости линии (L), направление которого согласовано с обходом (L) по правилу винта (п VII.11).

Аналогично можно было бы доказать формулу

$$\int_{(L)} x\,dy = S\cos(\widehat{\mathbf{n}, x}). \tag{91}$$

Если совершить циклическую перестановку (п. VII.12) координатных осей, то из формул (90) и (91) получим также

$$\left.\begin{aligned} &\int_{(L)} z\,dy = -S\cos(\widehat{\mathbf{n}, x}), && \int_{(L)} x\,dz = -S\cos(\widehat{\mathbf{n}, y}), \\ &\int_{(L)} y\,dz = S\cos(\widehat{\mathbf{n}, x}), && \int_{(L)} z\,dx = S\cos(\widehat{\mathbf{n}, y}). \end{aligned}\right\} \tag{92}$$

Отметим в заключение, что интегралы вида

$$\oint_{(L)} f(x)\,dx, \quad \oint_{(L)} \varphi(y)\,dy, \quad \oint_{(L)} \psi(z)\,dz$$

по *замкнутой* линии (L) (криволинейный интеграл по замкнутому контуру принято обозначать знаком $\oint$) всегда равны нулю. В самом деле, если $F(x)$ — первообразная к $f(x)$, то первый из интегралов равен приращению функции $F(x)$, когда точка обходит (L) и возвращается в исходное положение, т. е. равен нулю.

24. Условия независимости криволинейного интеграла по координатам от контура интегрирования. Рассмотрим интеграл вида

$$I=\int\limits_{(L)}[P(x,\ y,\ z)\,dx+Q(x,\ y,\ z)\,dy+R(x,\ y,\ z)\,dz], \tag{93}$$

где P, Q и R — некоторые функции, заданные во всем пространстве x, y, z или в некоторой его области и не обращающиеся там в бесконечность, а (L) — произвольная линия в этой области. В физических задачах иногда бывает, что интеграл (93) зависит только от положения начальной и конечной точек линии (L), но не зависит от того, как именно линия (L) проходит между этими точками. Другими словами (рис. 274),

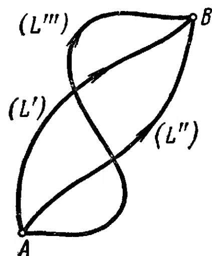


Рис. 274.

$$\int\limits_{(L')}(P\,dx+Q\,dy+R\,dz)=\int\limits_{(L'')}(P\,dx+Q\,dy+R\,dz)=\int\limits_{(L''')}(P\,dx+Q\,dy+R\,dz)=\ldots \tag{94}$$

при любом расположении точек A и B. *Тогда мы будем говорить, что интеграл* (93) *не зависит от контура интегрирования.* Например, интеграл (93) может иметь смысл работы силового поля при перемещении точки; тогда условие (94) означает, что эта работа зависит только от начального и конечного положений точки.

Для независимости интеграла (93) *от контура интегрирования необходимо и достаточно, чтобы интеграл* (93) *по любому замкнутому контуру равнялся нулю*, т. е. чтобы

$$\oint\limits_{(L)}(P\,dx+Q\,dy+R\,dz)=0 \tag{95}$$

для любого замкнутого контура (L).

Для доказательства предположим, что условие (95) выполнено и даны контуры (L') и (L'') с одинаковыми начальной и конечной точками (см. рис. 274). Построим замкнутый контур (L), идущий из A в B по (L') и из B в A по (L''), причем (L'') проходится в противоположном направлении. Применим условие (95), разобьем интег-

рал (95) на два в силу свойства 3 п. 4 и изменим у второго интеграла направление интегрирования в силу свойства 1 п. 4. Тогда получим

$$0=\oint_{(L)}=\int_{(L')}-\int_{(L'')},\ \text{т. е.}\ \int_{(L')}=\int_{(L'')},$$

откуда и вытекает (94). Рассуждая в обратном порядке, легко из (94) вывести (95).

Для независимости интеграла (93) *от контура интегрирования необходимо и достаточно, чтобы подынтегральное выражение было полным дифференциалом некоторой* (*однозначной!*) *функции трех переменных*, т. е. чтобы

$$P\,dx+Q\,dy+R\,dz\equiv du \tag{96}$$

для некоторой функции $u=u(x,\ y,\ z)$. Для доказательства предположим сначала, что условие (96) выполнено. Тогда

$$\int_{(L)}(P\,dx+Q\,dy+R\,dz)=\int_{(L)}du=u\,(B)-u\,(A),$$

где A и B — начало и конец линии (L). Значит, интеграл не зависит от контура интегрирования.

Пусть, обратно, интеграл (93) не зависит от контура интегрирования. Зафиксируем произвольно точку M_0 в пространстве и для любой текущей точки $M(x;\ y;\ z)$ определим значение функции

$$u\,(M)=\int_{\smile M_0M}(P\,dx+Q\,dy+R\,dz), \tag{97}$$

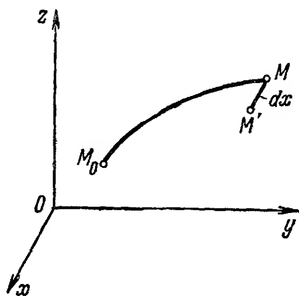


Рис. 275.

где $\smile M_0M$ — произвольная линия, идущая из M_0 в M. По условию эта функция однозначная, т. е. принимает в каждой точке M вполне определенное значение. Чтобы найти du, придадим сначала x бесконечно малое приращение dx. Тогда точка переместится в положение M' (рис. 275), причем бесконечно малый отрезок MM' будет параллелен оси x. Соответствующее приращение функции u равно

$$\Delta_x u=u\,(M')-u\,(M)=\int_{\smile M_0MM'}(P\,dx+Q\,dy+R\,dz)-$$
$$-\int_{\smile M_0M}(P\,dx+Q\,dy+R\,dz)=\int_{MM'}(P\,dx+Q\,dy+R\,dz).$$

Но так как вдоль отрезка MM' координаты y и z не меняются, то в последнем интеграле $dy=dz=0$ и $\Delta_x u=\int\limits_{MM'} P\,dx$.

Так как отрезок MM' бесконечно малый, то с точностью до малых высшего порядка можно считать на нем $P=\text{const}$, откуда, переходя от приращения к дифференциалу и тем самым отбрасывая малые высшего порядка, получим $\partial_x u=P\,dx$. Аналогично проверяем, что $\partial_y u=Q\,dy$, $\partial_z u=R\,dz$, и, складывая, получаем полный дифференциал (п. IX.11)

$$du=P\,dx+Q\,dy+R\,dz,$$

т. е. условие (96) выполнено.

Если вспомнить выражение (IX.7) для полного дифференциала, то условие (96) можно записать еще так:

$$\frac{\partial u}{\partial x}=P, \qquad \frac{\partial u}{\partial y}=Q, \qquad \frac{\partial u}{\partial z}=R. \tag{98}$$

Отсюда легко вывести, что если *интеграл* (93) *не зависит от контура интегрирования*, то

$$\frac{\partial P}{\partial y}\equiv\frac{\partial Q}{\partial x}, \qquad \frac{\partial P}{\partial z}\equiv\frac{\partial R}{\partial x}, \qquad \frac{\partial Q}{\partial z}\equiv\frac{\partial R}{\partial y}. \tag{99}$$

Действительно, из условий (98) получаем

$$\frac{\partial P}{\partial y}\equiv\frac{\partial^2 u}{\partial x\,\partial y}, \qquad \frac{\partial Q}{\partial x}\equiv\frac{\partial^2 u}{\partial y\,\partial x},$$

и в силу независимости смешанной производной от порядка дифференцирования (п. IX.15) получаем первое равенство (99); аналогично доказываются остальные (проверьте!).

В теории векторного поля (п. XVI.27) будет доказано обратное предложение: *если условия* (99) *выполнены и область, в которой рассматриваются функции* P, Q, R, *односвязна, то интеграл* (93) *не зависит от контура интегрирования*. При этом *односвязной* называется такая область, для которой любой расположенный в ней замкнутый контур можно путем непрерывной деформации стянуть в точку, не выходя за пределы области.

Например, все пространство, полупространство, двугранный или многогранный угол, внутренность или внешность сферы, внутренность конечного или бесконечного кругового цилиндров — все это односвязные области. В отличие от этого внешность бесконечного кругового цилиндра — это *неодносвязная* (двусвязная) область, так как, например, контур (L), изображенный на рис. 276, нельзя путем непрерывной деформации стянуть в точку, не выходя за пределы области. Неодносвязными являются также внутренность или внешность тора, а также полное пространство, из которого выброшены

все точки бесконечной прямой линии или окружности. На рис. 277 изображены примеры односвязной и четырехсвязной областей на плоскости. Плоскость становится двусвязной, даже если из нее выбросить одну лишь точку.

Сравнивая условия (96) и (99), мы видим, что в односвязной области условия (99) необходимы и достаточны для того, чтобы выражение $P\,dx+Q\,dy+R\,dz$ было полным дифференциалом некоторой однозначной функции $u(x, y, z)$. Можно показать, что

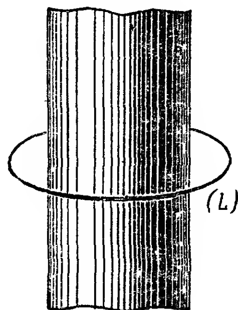
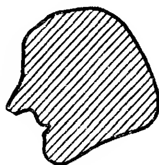


Рис. 276.



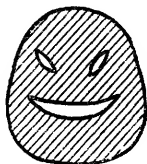


Рис. 277.

если условия (99) выполнены в многосвязной области, то функция u, построенная по формуле (97), удовлетворяет соотношению (96), но будет, вообще говоря, многозначной.

§ 7. Понятие об обобщенных функциях

25. Дельта-функция. *Дельта-фучкция* $\delta(x)$, широко применяемая в математике и ее приложениях, является простейшим примером *обобщенных функций*, теория которых заложена в 1936 г. советским математиком и механиком С. Л Соболевым и детально разработана лишь в последние пятнадцать лет.

Чтобы приближенно представить себе дельта-функцию, рассмотрим сначала разрывную функцию, определенную равенствами

$$y=\delta_N(x)=\begin{cases} 0 & \left(-\infty<x<-\frac{1}{2N}\right), \\ N & \left(-\frac{1}{2N}<x<\frac{1}{2N}\right), \\ 0 & \left(\frac{1}{2N}<x<\infty\right) \end{cases} \qquad (100)$$

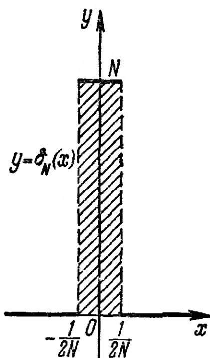


Рис. 278.

при очень большом N, ее график изображен на рис. 278. Значения функции в самих точках разрыва $x=\pm\frac{1}{2N}$ здесь, как обычно, несущественны. Дельта-функция получается в пределе, когда $N\to\infty$. Она, строго говоря, не имеет графика и может быть

определена так:

$$\delta(x)=\begin{cases}0 & (-\infty<x<-0;\ +0<x<\infty),\\ \infty & (-0<x<+0),\end{cases}$$

$$\int_{-\infty}^{\infty}\delta(x)\,dx=1;$$

впрочем, последнее равенство можно также записать в виде

$$\int_{-0}^{+0}\delta(x)\,dx=1. \qquad (101)$$

Для приближенного изображения дельта-функции не обязательно пользоваться разрывной функцией (100), можно положить, например, $\delta_N(x)=\frac{N}{\pi}(1+N^2x^2)^{-2}$ ($-\infty<x<\infty$; проследите за графиком этой функции при $N\to\infty$) и т. п. Вообще, годится любая функция, концентрирующая свои значения около $x=0$; более точно, достаточно только, чтобы $\delta_N(x)\geqslant 0$ ($-\infty<x<\infty$) и чтобы при $N\to\infty$ было $\int_a^b\delta_N(x)\,dx\to 0$, $\int_{-a}^{-b}\delta_N(x)\,dx\to 0$, $\int_{-a}^{b}\delta_N(x)\,dx\to 1$ для любых постоянных положительных чисел a и b.

Если рассматривать массы, распределенные вдоль оси x, и их линейные плотности, то окажется, что плотность единичной точечной массы, расположенной в начале координат, как раз равна дельта-функции. Действительно, если сначала представить себе эту массу не точечной, а равномерно распределенной на отрезке $-\frac{1}{2N}\leqslant x\leqslant\frac{1}{2N}$, то плотность будет иметь вид, изображенный на рис. 278 (почему?). Если теперь $N\to\infty$, то в пределе масса станет точечной, а плотность — дельта-функцией.

Аналогичным образом функция $\varrho(x)=m\delta(x-a)$ представляет собой плотность линейной (т. е. расположенной вдоль линии, в данном случае оси x) массы m, сосредоточенной в точке a. Так же можно изобразить плотности точечного заряда, сосредоточенной нагрузки и т. п.

Иногда приходится складывать дельта-функцию и обычную функцию. Например, сумма

$$\rho(x)=\rho_0+m_1\delta(x-a_1)+m_2\delta(x-a_2)$$

представляет собой плотность комбинации равномерно распределенной массы и двух точечных масс. Поэтому применение дельта-функции дает возможность во всех случаях распределенных, точечных и комбинированных масс применять формулы, относящиеся к распределенным массам; более того, само противопоставление распределенных масс точечным в значительной степени теряет смысл. То же относится к зарядам, нагрузкам и т. п.

При интегрировании выражений, содержащих дельта-функцию, надо иметь в виду формулу (101). Например, если $f(x)$ — непрерывная функция и $\alpha<a<\beta$, то

$$\int_\alpha^\beta f(x)\delta(x-a)\,dx=\int_\alpha^{a-0}f(x)\delta(x-a)\,dx+\int_{a-0}^{a+0}f(x)\delta(x-a)\,dx+$$
$$+\int_{a+0}^{\beta}f(x)\delta(x-a)\,dx=0+\int_{a-0}^{a+0}f(a)\delta(x-a)\,dx+0=f(a)\cdot 1=f(a); \quad (102)$$

при этом первый и третий интегралы равны нулю, так как там дельта-функция равна нулю, а во втором интеграле мы заменили $f(x)$ на $f(a)$, так как на бесконечно малом интервале непрерывную функцию можно считать постоянной, и воспользовались формулой (101).

Отметим двусмысленность интеграла вида $\int\limits_{\alpha}^{\beta} f(x)\,\delta(x-\alpha)\,dx$. Здесь обязательно надо указать, захватывается или нет при интегрировании особенность дельта-функции, так как $\int\limits_{\alpha-0}^{\beta} f(x)\,\delta(x-\alpha)\,dx=f(\alpha)$, $\int\limits_{\alpha+0}^{\beta} f(x)\,\delta(x-\alpha)\,dx=0$.

При интегрировании дельта-функции получается *единичная функция*

$$e(x)=\int\limits_{-\infty}^{x}\delta(t)\,dt=\begin{cases}0 & (-\infty<x<0),\\ 1 & (0<x<\infty),\end{cases} \qquad (103)$$

которая также имеет многочисленные применения. Ее график показан на рис. 279; она получается при внезапном подключении какого-либо постоянного воздействия, например напряжения в электрическую цепь.

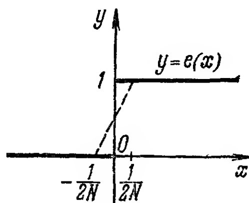


Рис. 279.

Итак, если проинтегрировать дельта-функцию, то получится обычная, не обобщенная функция, хотя и разрывная. При вторичном интегрировании получилась бы даже непрерывная функция (какая?).

Если продифференцировать равенство (103), то получится, что

$$\delta(x)=e'(x). \qquad (104)$$

Это равенство понимается в обобщенном смысле. Например, можно заменить на рис. 279 вертикальный отрезок на косой, соединяющий точки $\left(-\frac{1}{2N};\,0\right)$ и $\left(\frac{1}{2N};\,1\right)$ и изображенный пунктиром. Тогда разрывная функция заменится на непрерывную, производная которой имеет график, изображенный на рис. 278. Если теперь перейти к пределу при $N\to\infty$, мы и получим соотношение (104).

Таким образом, дельта-функция получается при дифференцировании разрывной функции, обладающей конечным скачком. Например, для закона движения, разобранного в п. 1.13 в связи с рис. 6, скорость выражается формулами

$$s'_t=\begin{cases}at & (0\leqslant t<t^*),\\ v & (t^*<t<\infty)\end{cases}$$

и потому при $t=t^*$ имеет скачок $s'_t(t^*+0)-s'_t(t^*-0)=at^*-v$. Поэтому ускорение равно $(at^*-v)\,\delta(t-t^*)+ae(t^*-t)$ (проверьте!). Первый член в этой сумме описывает явление *удара*.

26. Приложение к построению функции влияния. Одним из важных приложений дельта-функции является построение *функции влияния*, называемой также *функцией Грина* по имени английского математика Дж. Грина (1793—1841). Рассмотрим сначала пример. Пусть исследуется *прогиб* $h(x)$ балки под действием внешней поперечной нагрузки, приложенной с плотностью $p(x)$ (см. рис. 280, где изображен график или, как говорят, *эпюр* внешней нагрузки). Будем считать нагрузки не слишком большими, так

чтобы можно было воспользоваться законом линейности: при сложении внешних нагрузок прогибы складываются.

Обозначим через $y = G(x;\ \xi)$ прогиб в точке x, полученный в результате приложения в точке ξ единичной нагрузки. Эта функция $G(x;\ \xi)$ и называется функцией влияния в рассматриваемом примере. Мы сейчас покажем, что если она известна, то легко найти прогиб и от воздействия произвольной нагрузки с плотностью $p(x)$.

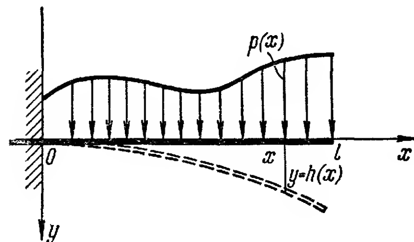


Рис. 280.

Действительно, рассмотрим нагрузку, приходящуюся на участок оси от точки ξ до точки $\xi + d\xi$. Эта нагрузка равна $p(\xi)\,d\xi$; поэтому прогиб от нее в точке x равен $G(x;\ \xi)\,p(\xi)\,d\xi$, так как из закона линейности вытекает, что если внешнюю нагрузку умножить на постоянный множитель, то и прогиб умножится на тот же множитель. Складываем все такие бесконечно малые прогибы, получаем суммарный прогиб (см. рис. 280):

$$h(x) = \int_0^l G(x;\ \xi)\,p(\xi)\,d\xi. \tag{105}$$

Перейдем теперь к общей схеме построения функции влияния. Пусть внешнее воздействие на какой-либо объект описывается функцией $f(x)$ ($a \leqslant x \leqslant b$; в приведенном примере это была функция $p(x)$), а результат этого воздействия — функцией $\tilde{f}(x)$ ($a \leqslant x \leqslant b$; в приведенном примере это была функция $h(x)$). Таким образом, каждая заданная функция f по какому-то определенному закону преобразуется в новую функцию $\tilde{f}$; в соответствии с п. XI.6 такой закон преобразования функций-*прообразов* в функции-*образы* называется *оператором*. Например, *оператор дифференцирования* D действует по закону $\mathrm{D}f = f'$, т. е. $\mathrm{D}(\sin x) = \cos x$, $\mathrm{D}(x^3) = 3x^2$ и т. д. Здесь $\sin x$ — прообраз, который преобразуется оператором D в образ $\cos x$, и т. д. Понятие оператора аналогично понятию функции (п. I.11), но если функция преобразовывала числа — значения независимой переменной — в числа — значения зависимой переменной, то оператор преобразует функции в функции.

Обозначим оператор перехода от функции внешнего воздействия $f(x)$ к функции «отклику» $\tilde{f}(x)$ через A, так что $\mathrm{A}f = \tilde{f}$. Мы предположим, что действует закон линейности, или, как говорят, *принцип суперпозиции: при сложении внешних воздействий их результаты также складываются.* Этот закон, который можно записать в форме

$$\mathrm{A}(f_1 + f_2) = \mathrm{A}f_1 + \mathrm{A}f_2,$$

часто применяется, когда внешние воздействия не слишком велики. Оператор, обладающий таким свойством, называется *линейным* (ср. п. XI 6). (Проверьте, что, например, оператор D линейный.) Из закона линейности можно вывести, что при умножении внешнего воздействия на константу его результат умножится на ту же константу, т. е.

$$\mathrm{A}(Cf) = C\mathrm{A}f \qquad (C = \mathrm{const}).$$

(Попробуйте это обосновать, считая сначала C целым положительным, затем полагая $C = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$, далее полагая $C = \frac{m}{n}$, где m и n — целые положительные, затем для $C = 0$ и, наконец, считая C отрицательным.)

Обозначим через $G(x;\ \xi)$ результат внешнего воздействия, описываемого дельта-функцией $\delta(x-\xi)$ при некотором фиксированном ξ, т. е.

$$\mathrm{A}[\delta(x-\xi)] = G(x;\ \xi).$$

Любую функцию $f(x)$ можно представить в виде суммы «столбчатых» функций (рис. 281), каждая из которых имеет особенность лишь в одной точке ξ, а потому равна $f(\xi)\,d\xi\,\delta(x-\xi)$ (почему?), т. е.

$$f(x) = \sum f(\xi)\,d\xi\,\delta(x-\xi).$$

Это, по существу, формула (102) в другой записи. Отсюда

$$\mathrm{A}[f(x)] = \mathrm{A}\left[\sum f(\xi)\,d\xi\,\delta(x-\xi)\right] = \sum \mathrm{A}[f(\xi)\,d\xi\,\delta(x-\xi)] = \sum f(\xi)\,d\xi\,\mathrm{A}[\delta(x-\xi)] = \sum f(\xi)\,d\xi\ G(x;\ \xi).$$

Но при бесконечно малых $d\xi$ эта сумма является интегралом, т. е. окончательно

$$\mathrm{A}[f(x)] = \int_a^b G(x;\ \xi)\,f(\xi)\,d\xi \quad (106)$$

(ср. с формулой (105)).

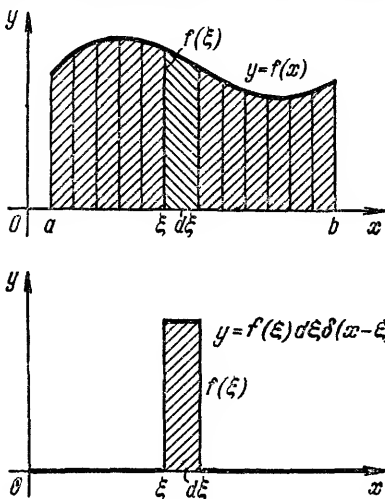


Рис. 281.

Функцию влияния можно в более простых случаях подсчитать теоретически (см., например, конец п. XV 16), а в более сложных определить экспериментально, производя необходимые замеры, например измеряя деформацию системы под действием сосредоточенной силы. Центральным пунктом является проверка *линейности системы*, т. е. возможности применения принципа суперпозиции; эта возможность выводится из общетеоретических принципов, либо проверяется экспериментально. Разумеется, не все системы являются точно или приближенно линейными. Грубую ошибку в этом отношении совершил господин де Графиньи, который, прогуливаясь в один летний день в двух черных камизолах, объяснил сей удивления достойный наряд так: «Вчера скончался дед мой, а сегодня испустила дух моя бабка! Для чего и надел я сугубый траур».

Отметим, что функции f и $\tilde{f}$ могут быть заданы на различных интервалах; более того, независимые переменные x и ξ в формуле (106) могут иметь разный физический смысл. Независимая переменная ξ может играть роль времени; тогда функция влияния описывает результат воздействия *единичного импульса*, подействовавшего в момент ξ.

Бывает, что рассматриваемая задача является линейной только при бесконечно малых внешних воздействиях, а дельтообразное внешнее воздействие нельзя считать малым. Тогда функцию влияния надо определить по формуле

$$G(x;\ \xi) = \lim_{P\to 0} \frac{1}{P}\,\mathrm{A}[P\delta(x-\xi)].$$

Так, в разобранном примере можно найти прогиб от малой нагрузки P, а результат разделить на P. В таких случаях формула (106) справедлива лишь для малых (точнее, бесконечно малых) по амплитуде внешних воздействий $f(x)$.

27. Другие обобщенные функции. Если при приближенном представлении дельта-функции (п. 25) воспользоваться непрерывной моделью, а затем произвести дифференцирование, то мы получим приближенное представление о $\delta'(x)$. У нее еще более «острая» особенность, чем у $\delta(x)$, причем $\delta'(x)$ принимает значения обоих знаков.

Если δ-функция описывает плотность единичного заряда, расположенного в начале координат (п. 25), то $\delta'(x)$ описывает плотность *диполя*, расположенного там же. В самом деле, такой диполь получится, если разместить заряды $-q$ и q соответственно в точках $x=0$ и $x=l$, а затем, оставляя $p=ql$ (*момент диполя*) без изменения, устремить l к нулю, так что в пределе получатся два равных бесконечно больших заряда противоположного знака на бесконечно близком расстоянии. До перехода к пределу плотность заряда имеет вид

$$q\delta(x-l)-q\delta(x)=-p\,\frac{\delta(x-l)-\delta(x)}{-l};$$

поэтому после перехода к пределу при $l\to 0$ плотность заряда равна $-p\delta'(x)$.

Интегралы с участием $\delta'(x)$ вычисляются с помощью интегрирования по частям: если $f(x)$ имеет непрерывную производную и $\alpha<a<\beta$, то

$$\int\limits_{\alpha}^{\beta} f(x)\,\delta'(x-a)\,dx=f(x)\,\delta(x-a)\Big|_{x=a}^{\beta}-\int\limits_{\alpha}^{\beta}\delta(x-a)\,f'(x)\,dx=-f'(a).$$

Интересно отметить, что вычисление этого интеграла по такому методу:

$$\int\limits_{\alpha}^{\beta} f(x)\,\delta'(x-a)\,dx=\int\limits_{a-0}^{a+0} f(x)\,\delta'(x-a)\,dx=f(a)\int\limits_{a-0}^{a+0}\delta'(x-a)\,dx=$$

$$=f(a)\,\delta(x-a)\Big|_{x=a-0}^{a+0}=0$$

является ошибочным, так как замена $f(x)$ на $f(a)$ при весьма острой особенности у $\delta'(x-a)$ является слишком грубой. Аналогично можно рассматривать и дальнейшие производные от дельта-функции, а также их комбинации.

Обобщенные функции можно классифицировать по числу интегрирований, после которых из заданной функции получится непрерывная функция. При такой классификации непрерывные функции можно считать обобщенными функциями нулевого порядка; функции с конечными разрывами и обычные функции с интегрируемыми особенностями на конечном расстоянии (п. 16) — обобщенными функциями первого порядка. Функция $\delta(x)$ — это простейший пример функции второго порядка (см. п. 25), $\delta'(x)$ имеет третий порядок и т. д. При дифференцировании функции ее порядок повышается на единицу, а при интегрировании — понижается на единицу.

При истолковании функции с неинтегрируемой особенностью в качестве обобщенной функции надо указывать, производной от какой функции нулевого или первого порядка она является. Например, функцию $\frac{1}{x}$, имеющую неинтегрируемую особенность при $x=0$, можно при $x\neq 0$ считать равной либо

$$(\ln|x|)', \qquad (107)$$

либо же, скажем,

$$(\ln|x|+e(x))', \qquad (108)$$

так как при $x\neq 0$ будет $e'(x)=0$. Однако функции (107) и (108) различаются на $e'(x)=\delta(x)$ и потому их свойства неодинаковы. Поэтому в теории обоб-

щенных функций вместо $\frac{1}{x}$ предпочитают применять запись (107) или (108) (что, конечно, не одно и то же), так как смысл такой записи уже однозначен. Аналогично функцию $\frac{1}{x^2}$ можно рассматривать как $-(\ln|x|)''$ и т. п. Функции, имеющие в своих точках разрыва рост выше степенно́го, например $e^{\frac{1}{|x|}}$, приходится при этом исключить; впрочем, практическое значение этих функций незначительное. Рост функции при $x \to \pm\infty$ не ограничивается.

Оказывается, что если пользоваться обобщенными функциями, то большинство правил, связанных с дифференцированием тех или иных формул, будет справедливым без каких-либо ограничений на характер участвующих функций. Например, правило Лейбница о дифференцировании по параметру под знаком интеграла (п 20) становится справедливым, независимо от характера сходимости рассматриваемых интегралов и т. п.

ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Дифференциальное уравнение — это уравнение, связывающее две или более функционально зависящих друг от друга величин и их дифференциалы или, что равносильно, производные. Задача о составлении и решении (как говорят, *интегрировании*) таких уравнений часто возникает в физике и технике.

§ 1. Общие понятия

1. Примеры. Простейшие дифференциальные уравнения уже встречались в нашем курсе. Например, уравнение (XIV.22) в случае силы, изменяющейся по закону Ньютона (п. XIV.14), можно переписать в виде

$$dA = \frac{k}{s^2}\,ds \quad \text{или} \quad \frac{dA}{ds} = \frac{k}{s^2}, \tag{1}$$

где работа $A = A(s)$ — неизвестная функция от перемещения, которая находится при помощи интегрирования.

Уравнение (XIV.27) можно переписать в виде

$$\frac{dh}{dt} = -\frac{\sigma}{S}\sqrt{2gh}, \tag{2}$$

где $h = h(t)$ — неизвестная функция.

Рассмотрим еще пример упругих колебаний материальной точки массы M около положения равновесия (рис. 282). Здесь неизвестным является закон колебаний $y = y(t)$. Примем для простоты *линейный закон упругости*, т. е. будем считать силу упругости прямо пропорциональной отклонению точки от положения равновесия; тогда эта сила равна

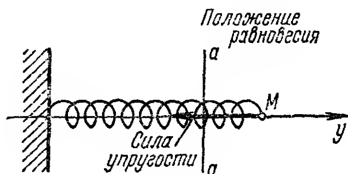


Рис. 282.

$$F = -ky,$$

где k — *коэффициент упругости*. Согласно второму закону Ньютона, если других сил нет, получим

$$M\frac{d^2y}{dt^2} = F = -ky, \tag{3}$$

откуда дифференциальное уравнение для закона колебаний приобретает вид

$$M\frac{d^2y}{dt^2}+ky=0, \qquad y=y(t)=? \tag{4}$$

Во всех случаях при выводе дифференциального уравнения в физике или технике берется за основу некоторый общий физический или какой-либо иной закон, имеющий *дифференциальный характер* (*дифференциальный закон*), т. е. связывающий бесконечно малые изменения рассматриваемых величин. После интегрирования уравнения получается *интегральный закон*, связывающий конечные значения этих величин. Вывод основных дифференциальных уравнений в той или иной области знания — очень ответственное дело, так как та или иная их формулировка в значительной степени определяет дальнейший ход развития этой области. Математическая зрелость физика или инженера во многом характеризуется тем, насколько правильно он может *математически формулировать* основные задачи в избранной им области.

2. Основные определения. Чаще всего *дифференциальное уравнение получается как уравнение, связывающее аргумент* или аргументы, *неизвестную функцию* (или функции) и *ее производные*; даже если первоначально было соотношение между дифференциалами, то можно перейти к соотношению между производными (см. формулы (1)). *Если искомая функция зависит от одного аргумента*, как в примерах п. 1, *то дифференциальное уравнение называется обыкновенным*; в противном случае оно называется *уравнением с частными производными* (почему?). *В этой главе будут рассматриваться только обыкновенные дифференциальные уравнения.*

Наивысший порядок производной от искомой функции, входящий в уравнение, называется порядком этого уравнения. Так, уравнения (1) и (2) — первого порядка, тогда как уравнение (4) имеет второй порядок. В общем случае уравнение n-го порядка имеет вид

$$F(x,\ y,\ y',\ y'',\ \ldots,\ y^{(n)})=0, \tag{5}$$

где $y=y(x)$ — искомая функция. Конечно, при этом функция F может фактически зависеть не от всех выписанных величин: например, уравнение (4) не включает независимую переменную и производную первого порядка.

Решением дифференциального уравнения называется функция, которая, будучи подставлена в это уравнение, обращает его в тождество. Уже на простейших примерах легко убедиться в том, что *дифференциальное уравнение имеет бесконечное количество решений.* Например, из простейшего уравнения

$$y'=x^2, \quad y=y(x) \tag{6}$$

сразу найдем с помощью интегрирования

$$y=\frac{x^3}{3}+C. \tag{7}$$

Это — *общее решение* уравнения (6); оно включает *произвольную постоянную* C и является записью *всего многообразия решений*. Придавая произвольной постоянной конкретные численные значения, мы получим конкретные, *частные решения* уравнения (6):

$$y=\frac{x^3}{3},\quad y=\frac{x^3}{3}+6,\quad y=\frac{x^3}{3}-\frac{\sqrt{2}}{3} \text{ и т. п.}$$

В общем случае (5) решение находится в результате n последовательных интегрирований, так что *общее решение уравнения n-го порядка содержит n произвольных постоянных*, т. е. имеет вид

$$y=y(x;\ C_1,\ C_2,\ \ldots,\ C_n). \tag{8}$$

Особенно часто общее решение получается в неявной форме:

$$\Phi(x,\ y;\ C_1,\ C_2,\ \ldots,\ C_n)=0. \tag{9}$$

Соотношения (8) и (9) называются также *общими интегралами* уравнения (5). Частные решения получаются, если придать каждой произвольной постоянной $C_1,\ C_2,\ \ldots,\ C_n$ конкретное численное значение. График каждого частного решения называется *интегральной линией рассматриваемого дифференциального уравнения*. Уравнение этой линии — это уравнение (8) и (9) с конкретными C.

Чтобы из общего решения выделить одно частное, требуется, помимо дифференциального уравнения, поставить некоторые дополнительные условия. Чаще всего ставятся *начальные условия*, которые при исследовании процесса, развивающегося во времени, являются *математической записью начального состояния процесса*.

Например, при рассмотрении процесса колебаний в п. 1 из физических соображений ясно, что конкретное колебание полностью определяется, если заданы начальное отклонение и начальная скорость колеблющейся точки. Поэтому начальные условия для уравнения (5) имеют вид

$$\text{при}\quad t=t_0\quad \text{заданы}\quad y=y_0 \ \text{и}\ \frac{dy}{dt}=v_0. \tag{10}$$

В общем случае для уравнения (5) начальные условия имеют следующий вид:

$$\text{при } x=x_0 \text{ заданы } y=y_0,\quad y'=(y')_0,\ \ldots,\ y^{(n-1)}=(y^{(n-1)})_0. \tag{11}$$

Так как общее решение (9) содержит n произвольных постоянных, то наложенных n соотношений как раз достаточно, во всяком случае принципиально, для нахождения этих постоянных и тем самым для нахождения частного решения. И физически естественно, что если известны дифференциальный закон, управляющий развитием процесса, а также начальное состояние этого процесса, то сам процесс является полностью определенным.

Для уравнения первого порядка (6) условие (11) означает, что при некотором значении $x=x_0$ должно быть задано значение $y=y_0$. Пусть, например, требуется, чтобы $y(1)=2$. Тогда из (7) получаем $2=\frac{1^3}{3}+C$, $C=\frac{5}{3}$, т. е. искомое частное решение имеет вид

$$y=\frac{x^3+5}{3}.$$

Задача о нахождении частного решения дифференциального уравнения при заданном начальном условии называется *задачей Коши*.

Как мы увидим в п. 7, некоторые дифференциальные уравнения, помимо частных решений, входящих в состав общего, обладают дополнительно *особыми* решениями, которые не входят в этот состав.

§ 2. Уравнения первого порядка

3. Геометрический смысл. Общий вид дифференциального уравнения первого порядка таков:

$$F(x,\ y,\ y')=0, \tag{12}$$

где $y=y(x)$ — неизвестная функция. Будем сначала для простоты считать, что уравнение разрешено относительно производной от искомой функции, т. е. имеет вид

$$y'=f(x,\ y). \tag{13}$$

Начальное условие для уравнения (12) согласно п. 2 таково:

$$\text{при } x=x_0 \text{ задано } y=y_0. \tag{14}$$

Чтобы придать геометрический смысл уравнению (13), рассмотрим плоскость x, y; тогда каждое частное решение изобразится в виде линии (интегральной линии) на этой плоскости, но эти линии нам пока неизвестны. Однако, взяв любую точку $M(x;\ y)$ на плоскости, можно подсчитать значение $f(x,\ y)$, которое в силу уравнения (13) должно равняться угловому коэффициенту касательной (п. IV.3) к искомой линии в точке M, если эта линия пройдет через M.

Поэтому можно поступить так: в каждой точке $M(x;\ y)$ плоскости представим себе проведенным маленький отрезок с угловым коэффициентом $\operatorname{tg}\alpha=f(x,\ y)$ (рис. 283). Конечно, мы на практике можем провести только несколько таких отрезков, но теоретически надо представлять себе, что отрезки проведены в *каждой* точке. Полученная картина называется *полем направлений* на плоскости (общее понятие поля см. в п. IX.9). Мы видим, что интегральные линии уравнения (13) должны проходить так, чтобы в каждой своей точке «идти вдоль поля», т. е. касаться отрезка, проведенного в этой точке.

Итак, *уравнение* (13) *задает на плоскости поле направлений.* С другой стороны, *начальное условие* (14) *задает на плоскости точку* $M_0(x_0;\ y_0)$, *через которую должна пройти искомая интегральная линия.* Геометрически ясно (рис. 284), что этим интегральная линия полностью определяется. Другими словами, *уравнение* (13) *при*

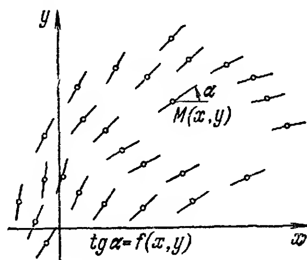


Рис. 283.

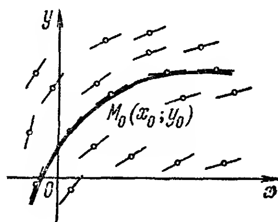


Рис. 284.

начальном условии (14) *имеет вполне определенное, единственное решение.* Более подробное исследование, проведенное Коши, показало, что для этого достаточно, чтобы в точке M_0 *функция* f *была непрерывна, а ее производная* $\frac{\partial f}{\partial y}$ — *конечна.* (Случаи нарушения условий этой *теоремы Коши* будут рассмотрены в п. 7.)

Приведенные рассуждения хорошо иллюстрируются на известном опыте с железными опилками, помещенными в магнитное поле. Сами опилки образуют поле направлений, а интегральными линиями этого поля служат так называемые магнитные силовые линии.

Выясненный нами геометрический смысл уравнения (13) дает возможность приближенно графически строить интегральные линии этого уравнения. Для этого надо изобразить поле направлений по возможности в большем числе точек плоскости, а затем строить линии, руководствуясь этими направлениями.

При построении поля на практике удобнее не выбирать произвольно точки на плоскости, а строить *изоклины*, т. е. линии, на которых поле направлено одинаково. Их уравнение получится, если правую часть уравнения (13) приравнять константе, т. е. написать

$$f(x,\ y)=k,$$

где k — угловой коэффициент поля, отвечающий данной изоклине.

Рассмотрим, например, уравнение $y'=x+y$.

Приравнивание правой части постоянным $-2,\ -\frac{3}{2},\ -1,\ -\frac{1}{2},\ 0,\ 1,\ 2$ дает изоклины, которые в данном примере являются прямыми линиями ($x+y=-2$ и т. д.), изображенными на рис. 285. На каждой из этих изоклин штрихами показано направление поля; чтобы найти это направление, мож-

но построить прямоугольный треугольник PQR с основанием $PQ=1$, параллельным оси x, и высотой $QR=k$, тогда сторона PR пойдет как раз по требуемому направлению. На основе этих направлений на рис. 285 нанесено также несколько интегральных линий. Видно, что одной из таких линий служит прямая $x+y=-1$, а также, что геометрическим местом самых низких точек на интегральных линиях служит прямая $x+y=0$. В общем случае (13) для нахождения геометрического места самых низких или самых высоких точек на интегральных линиях надо построить изоклину $f(x, y)=0$. (Как в общем случае найти геометрическое место точек перегиба интегральных линий?)

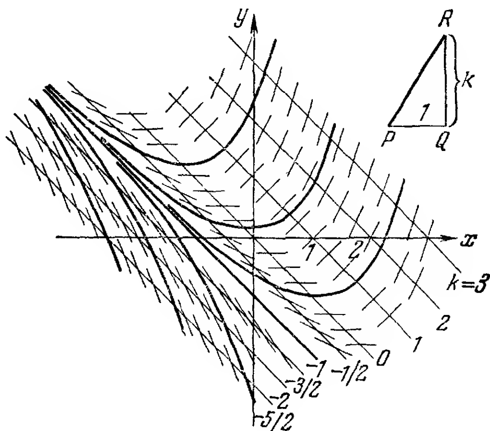


Рис. 285.
PR — направление поля.

4. Интегрируемые типы уравнений. *Дифференциальное уравнение считается проинтегрированным в квадратурах, если его общее решение получено в явной или неявной форме, которая может содержать еще не взятые интегралы от известных функций*; как быть с такими интегралами — учит интегральное исчисление (гл. XIII и XIV). К сожалению, многие даже очень простые уравнения невозможно проинтегрировать в квадратурах и их приходится исследовать иными методами, о которых в свое время будет сказано. Тем не менее имеется несколько классов уравнений, интегрируемых в квадратурах.

1. *Дифференциальные уравнения с разделяющимися переменными* уже были рассмотрены в п. XIV.7; они имеют общий вид

$$\frac{dy}{dx}=f(x)\varphi(y) \tag{15}$$

и общее решение

$$\int\frac{dy}{\varphi(y)}=\int f(x)\,dx+C. \tag{16}$$

Мы написали в правой части произвольную постоянную C, чтобы подчеркнуть, что она входит в общее решение. Как видим, уравнение (15) проинтегрировано в квадратурах формулой (16).

Рассмотрим, например, простейшее уравнение

$$\frac{dy}{dx}=2xy.$$

Разделяя переменные и интегрируя, получим

$$\frac{dy}{y}=2x\,dx,\quad \ln|y|=x^2+C,\quad |y|=e^{x^2+C},\quad y=\pm e^Ce^{x^2}. \tag{17}$$

Ответ удобнее записать в иной форме, если заметить, что $\pm e^C$ также представляет собой произвольную постоянную, т. е. $y=Ce^{x^2}$; конечно, здесь C уже иное, чем в формуле (17).

Чтобы избежать такого изменения обозначений, можно было при интегрировании в (17) написать, имея в виду дальнейшее потенцирование,

$$\ln|y|=x^2+\ln C,$$

так как $\ln C$—тоже произвольная постоянная После потенцирования получим

$$|y|=Ce^{x^2},\quad y=\pm Ce^{x^2},\quad y=Ce^{x^2};$$

знак можно включить в C, т. е. в последней формуле считать C любого знака Мы будем так поступать без особого предупреждения.

Аналогично (16) строится общее решение уравнения

$$P(x)\,Q(y)\,dx+R(x)\,S(y)\,dy=0,$$

которое также является уравнением с разделяющимися переменными.

2. *Уравнение, однородное относительно аргумента и искомой функции.* Так называется уравнение вида (12), если его левая часть представляет собой однородную функцию относительно x и y, рассматриваемых как независимые переменные, т. е. если

$$F(tx,\ ty,\ y')\equiv t^kF(x,\ y,\ y')$$

(общее определение однородной функции было дано в п. IX.12).

Тогда уравнение (12) можно переписать в виде

$$F\left(x\cdot 1,\ x\cdot\frac{y}{x},\ y'\right)=0,\quad x^kF\left(1,\ \frac{y}{x},\ y'\right)=0,\quad \text{т. е.}\quad F\left(1,\ \frac{y}{x},\ y'\right)=0.$$

Разрешая последнее уравнение относительно y', получим

$$y'=\varphi\left(\frac{y}{x}\right). \tag{18}$$

Это уравнение интегрируется с помощью подстановки

$$\frac{y}{x}=u,\quad y=ux,\quad y'=u'x+u,$$

где $u=u(x)$—новая, взамен y, искомая функция. Подставляя в (18), получим

$$u'x+u=\varphi(u),\quad \frac{du}{dx}x=\varphi(u)-u,\quad \frac{du}{\varphi(u)-u}=\frac{dx}{x},$$

т. е. переменные разделились; теперь надо выполнить интегрирование и перейти от u обратно к y.

Для интегрирования несколько более общего, чем (18), уравнения

$$\frac{dy}{dx}=\varphi\left(\frac{ax+by+c}{mx+ny+p}\right),$$

где двучлены $ax+by$ и $mx+ny$ не пропорциональны друг другу, совершают подстановку $x=x_1+\alpha$, $y=y_1+\beta$, где α и β подбираются так, чтобы после подстановки в числителе и знаменателе последней дроби не осталось свободных членов; затем заменяют $y_1/x_1=u$.

3. *Линейные уравнения.* Так называются уравнения вида (12), *линейные относительно искомой функции и ее производной*, т. е. уравнения вида

$$a(x)y'+b(x)y+c(x)=0.$$

Если это уравнение разделить на $a(x)$, получится

$$y'+p(x)y=f(x), \quad \text{где } p=\frac{b}{a}, \; f=-\frac{c}{a}. \tag{19}$$

Уравнение (19) называется *линейным неоднородным* из-за наличия в нем неоднородного члена $f(x)$. Чтобы решить это уравнение, напишем сначала вспомогательное *однородное уравнение*, отбросив неоднородный член и изменив обозначение искомой функции:

$$z'+p(x)z=0. \tag{20}$$

В уравнении (20) переменные разделяются:

$$\frac{dz}{dx}=-p(x)z, \quad \frac{dz}{z}=-p(x)\,dx, \quad \ln|z|=-\int p(x)\,dx+\ln C,$$

$$z=C\exp\left[-\int p(x)\,dx\right]=Cz_1, \tag{21}$$

где z_1 получается из z при $C=1$, т. е. z_1 есть частное решение уравнения (20).

Найдя z, ищем решение уравнения (19) в виде

$$y=\varphi(x)z_1, \tag{22}$$

где z_1 то же, что в формуле (21), а $\varphi(x)$ — пока неизвестная функция. Такая замена бывшей произвольной постоянной из формулы (21) на функцию в формуле (22) называется *вариацией произвольной постоянной*.

Из формулы (22) получим, подставляя в уравнение (19),

$$\varphi' z_1+\varphi z_1'+p\varphi z_1=f, \quad \varphi' z_1+\varphi(z_1'+pz_1)=f.$$

Так как z_1 удовлетворяет уравнению (20), то последняя скобка равна нулю, т. е.

$$\varphi'(x)=\frac{f(x)}{z_1(x)}, \quad \varphi(x)=\int\frac{f(x)}{z_1(x)}\,dx+C,$$

$$y=z_1(x)\int\frac{f(x)}{z_1(x)}\,dx+Cz_1(x).$$

Это — общее решение уравнения (19). Первое слагаемое получается при $C=0$ и, таким образом, представляет собой частное решение того же уравнения.

Итак, *общее решение линейного неоднородного уравнения* (19) *равно сумме некоторого его частного решения и общего решения соответствующего однородного уравнения* (20), *полученного отбрасыванием неоднородного члена*. В п. 15 мы увидим, что этим свойством обладают линейные уравнения любого вида.

Уравнение

$$y'+p(x)\,y=f(x)\,y^n$$

называется *уравнением И. Бернулли*. Оно приводится к линейному, если разделить обе части на y^n и сделать замену $y^{1-n}=u$ (проверьте!).

Имеется еще несколько мало распространенных типов уравнений, интегрируемых в квадратурах. Кроме того, отдельные уравнения, не принадлежащие к определенным типам, иногда удается проинтегрировать в результате удачно подобранной подстановки. Значительная часть таких уравнений приведена в книге [22]. В то же время большинство уравнений не интегрируется в квадратурах. Например, не интегрируется в общем случае довольно распространенное *уравнение Риккати* (Ф. Рикка́ти (1676—1754) — итальянский математик)

$$y'=y^2+f(x)$$

и многие другие простые уравнения.

5. Уравнение для экспоненты. Остановимся на очень простом, но крайне важном уравнении

$$\frac{dy}{dx}=ky \qquad (k=\text{const}) \tag{23}$$

Оно означает, что скорость изменения величины y, взятая по отношению к величине x, пропорциональна текущему значению y. Такая пропорциональность с $k>0$, если y возрастает, и с $k<0$, если y убывает (мы для простоты считаем, что $y>0$), часто принимается в первом приближении при исследовании многих процессов, а иногда она оправдывается с большой точностью

В уравнении (23) разделяются переменные, откуда

$$\frac{dy}{y}=k\,dx, \quad \ln|y|=kx+\ln C, \quad y=Ce^{kx}.$$

Если имеется также начальное условие $y(x_0)=y_0$, то получаем

$$y_0=Ce^{kx_0}, \quad C=y_0e^{-kx_0}, \quad \text{т. е.} \quad y=y_0e^{k(x-x_0)}. \tag{24}$$

Итак, решение уравнения (23) представляет собой экспоненту, т. е. показательную функцию (п. I 27). Для решения характерно, что если придавать x значения, образующие арифметическую прогрессию с разностью Δx, то соответствующие значения y образуют геометрическую прогрессию со знаменателем $e^{k\Delta x}$ Легко найти, каково должно быть Δx, чтобы y менялся (увеличивался или уменьшался) каждый раз вдвое. Для этого должно быть

$$|k\,\Delta x|=\ln 2, \qquad \text{т. е.} \qquad \Delta x=\frac{\ln 2}{|k|}. \tag{25}$$

Если $k>0$, то формула (25) показывает *экспоненциальное нарастание* величины y. Так получится, например, при исследовании процесса размножения бактерий в питательной среде, пока их там не слишком много. При-

мем, что все они размножаются более или менее независимо друг от друга; это — так называемый *закон органического роста*, характерный для всевозможных *цепных реакций* Тогда получаем, что скорость нарастания количества u этих бактерий, измеренного в каких-то единицах, пропорциональна этому количеству, т. е.

$$\frac{du}{dt}=ku; \qquad u=u_0e^{k(t-t_0)}.$$

Аналогично исследуются задача о непрерывном нарастании вклада в сберкассе и другие подобные задачи

Если $k<0$, то формула (24) показывает *экспоненциальное убывание* величины y Так получится, например, при исследовании процесса радиоактивного распада. Если принять, что различные участки распадаются независимо друг от друга, то получаем, что скорость убывания еще не распавшейся массы m радиоактивного вещества пропорциональна текущему значению этой массы, т. е.

$$\frac{dm}{dt}=-pm, \qquad m=m_0e^{-p(t-t_0)}.$$

Отметим, в частности, что в силу формулы (25) за время $\Delta t=\frac{\ln 2}{p}$ значение m уменьшается наполовину; это — *период полураспада*. Так, для радия он приближенно равен $1{,}8\cdot 10^3$ лет, другими словами, если бы запасы радия не пополнялись, то через $1{,}8\cdot 10^3$ лет осталась бы половина начального запаса, еще через $1{,}8\cdot 10^3$ лет — четверть начального запаса и т. д.

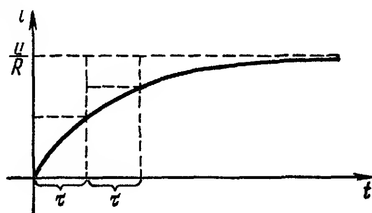


Рис. 286.

Аналогично исследуются убывание атмосферного давления с высотой, процесс разрядки конденсатора через сопротивление и многие другие задачи

Иногда рассматриваемое уравнение можно более или менее просто преобразовать к виду (23) Например, согласно п. VIII.7 при включении постоянного напряжения u в цепь, обладающую сопротивлением R и индуктивностью L, ток i удовлетворяет уравнению

$$L\frac{di}{dt}+Ri=u \tag{26}$$

Это — линейное неоднородное уравнение, которое можно проинтегрировать (решить) по методу п. 4. Но проще преобразовать уравнение так:

$$L\frac{di}{dt}=-Ri+u=-R\left(i-\frac{u}{R}\right), \qquad \frac{d\left(i-\frac{u}{R}\right)}{dt}=-\frac{R}{L}\left(i-\frac{u}{R}\right),$$

откуда

$$i-\frac{u}{R}=\left(i_0-\frac{u}{R}\right)e^{-\frac{R}{L}(t-t_0)}, \qquad i=\frac{u}{R}+\left(i_0-\frac{u}{R}\right)e^{-\frac{R}{L}(t-t_0)}.$$

Особенно просто получится, если в начальный момент, за который мы примем $t=0$, тока в цепи не было. Тогда $t_0=0$, $i_0=0$ и

$$i=\frac{u}{R}\left(1-e^{-\frac{R}{L}t}\right). \tag{27}$$

График полученной зависимости показан на рис. 286. Мы видим, что ток при $t\to\infty$ экспоненциально приближается к предельному стационарному

значению $\frac{u}{R}$. Это же значение легко найти из самого уравнения (26), если заметить, что в процессе установления тока, при $t \to \infty$, будет $\frac{di}{dt} \to 0$, и потому в пределе $Ri = u$, $i = \frac{u}{R}$, т. е. когда ток практически установился, все напряжение расходуется только на сопротивление R. Отклонение тока от предельного значения уменьшается в два раза за время

$$\tau = \frac{\ln 2}{\frac{R}{L}} = \frac{L}{R} \ln 2.$$

То, что в формуле (24) в основании получается именно число e, и есть основная причина значения этой константы в математике и ее приложениях.

6. Интегрирование полного дифференциала. Дифференциальные уравнения первого порядка часто рассматривают взамен формы (13) в *симметричной форме*

$$P(x, y)\,dx + Q(x, y)\,dy = 0, \tag{28}$$

где $P(x, y)$ и $Q(x, y)$ — заданные функции, а функциональная зависимость между x и y неизвестна. Легко перейти от одной формы к другой: например, чтобы перейти от (28) к форме (13), надо обе части (28) разделить на $Q\,dx$, а затем перенести $\frac{P}{Q}$ в правую часть. Форма (28) предпочтительнее в тех случаях, когда переменные x и y более или менее равноправны и заранее не требуется, чтобы именно y считался функцией x, а не наоборот.

В частном случае, когда левая часть уравнения (28) представляет собой полный дифференциал некоторой функции, т. е.

$$P\,dx + Q\,dy \equiv du(x, y), \tag{29}$$

это уравнение (28) легко проинтегрировать. Действительно, тогда его можно переписать в виде $du = 0$ и, интегрируя, получим общее решение

$$u(x, y) = C, \tag{30}$$

где C — как всегда, произвольная постоянная.

Вопрос о том, когда существует такая функция $u(x, y)$, был рассмотрен в п. XIV.24. В данном случае зависимости от z нет и $R \equiv 0$, так что из условий (XIV.99) остается только одно:

$$\frac{\partial P}{\partial y} \equiv \frac{\partial Q}{\partial x}. \tag{31}$$

Это условие необходимо и достаточно для того, чтобы левая часть уравнения (28) *была полным дифференциалом.* При этом функция $u(x, y)$ строится по формуле (XIV.97), где, конечно, под знаком интеграла должно быть отброшено последнее слагаемое. Как указа-

но в п. XIV.24, если область многосвязная, то функция u получится, вообще говоря, многозначной. Однако формула (30) и в этом случае дает общее решение уравнения (28).

Рассмотрим, например, уравнение

$$(x^2+2xy)\,dx+(x^2-y^3)\,dy=0. \tag{32}$$

Здесь

$$\frac{\partial P}{\partial y}=\frac{\partial (x^2+2xy)}{\partial y}=2x, \qquad \frac{\partial Q}{\partial x}=\frac{\partial (x^2-y^3)}{\partial x}=2x,$$

т. е. условие (31) выполнено. Для построения функции u по формуле (XIV.97) выберем для определенности точку M_0 в начале координат, а путь, соединяющий M_0 с текущей точкой $M(x;\ y)$, — как на рис. 287. Получаем

$$u(x,\ y)=\int\limits_{M_0M}[(x^2+2xy)\,dx+(x^2-y^3)\,dy]=$$

$$=\int\limits_{M_0M'}[(x^2+2xy)\,dx+(x^2-y^3)\,dy]+$$

$$+\int\limits_{M'M}[(x^2+2xy)\,dx+(x^2-y^3)\,dy]. \tag{33}$$

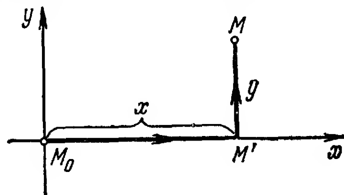


Рис. 287.

В первом интеграле надо положить $y=0$, $dy=0$, тогда как во втором считать $x=\text{const}$, $dx=0$ (почему?). Отсюда

$$u(x,\ y)=\int\limits_0^x x^2\,dx+\int\limits_0^y (x^2-y^3)\,dy=\frac{x^3}{3}+x^2y-\frac{y^4}{4}. \tag{34}$$

Итак, общее решение уравнения (32) имеет вид

$$\frac{x^3}{3}+x^2y-\frac{y^4}{4}=C.$$

Проведем аналогичное исследование уравнения

$$-\frac{y\,dx}{x^2+y^2}+\frac{x\,dy}{x^2+y^2}=0, \tag{35}$$

которое, впрочем, легко проинтегрировать непосредственно, так как переменные в нем разделяются. Это уравнение приходится рассматривать на всей плоскости, за исключением начала координат, как говорят, на плоскости с *выколотым* началом координат, так как при $x=0$, $y=0$ оба коэффициента, P и Q, имеют разрыв. Такая область, как указано в п. XIV.24, неодносвязная (двусвязная).

Для уравнения (35) условие (31) также выполняется (проверьте!). Для построения функции u по формуле (XIV.97) выберем точку M_0 где угодно, но, конечно, не в начале координат, например $M_0(1;\ 0)$. Проведя выкладки, аналогичные (33), (34), и считая сначала, что $x>0$, получим (проверьте!) $u=\operatorname{arctg}\frac{y}{x}$. Эта же функция удовлетворяет соотношению (29) и при $x<0$; однако если ее рассмотреть во всей плоскости x, y, то она будет иметь разрыв на прямой $x=0$. Чтобы избавиться от него, можно положить

$$u=\operatorname{Arctg}\frac{y}{x}=\varphi \text{ (полярному углу).}$$

Правда, эта функция неоднозначна: даже если в некоторой точке $M \neq O$ выбрать какое-либо одно значение φ, а затем заставить M обойти вокруг начала координат, то φ получит приращение 2π. Тем не менее общее решение уравнения (35) имеет вид

$$\operatorname{Arctg}\frac{y}{x}=C, \quad \text{т. е.} \quad \frac{y}{x}=\operatorname{tg} C=C_1, \quad y=C_1x,$$

где C_1 — произвольная постоянная; геометрически получаем семейство всевозможных прямых, проходящих через начало координат.

Бывает так, что для уравнения (28) условие (31) не выполнено, т. е. левая часть этого уравнения не является полным дифференциалом, но становится им после умножения на некоторый известный множитель. Например, левая часть уравнения $-y\,dx+x\,dy=0$ не удовлетворяет условию (31), но начинает удовлетворять после умножения обеих частей на множитель $\frac{1}{x^2+y^2}$ (см. (35)). Такой множитель называется *интегрирующим множителем* для рассматриваемого уравнения (28). Никаких общих способов для его нахождения нет; интегрирующий множитель используется в некоторых теоретических исследованиях.

7. Особые точки и особые решения. Бывает, что для уравнения, записанного в форме

$$y'=f(x, y) \tag{36}$$

(см. уравнение (13)) или

$$P(x, y)\,dx+Q(x, y)\,dy=0 \tag{37}$$

(см. уравнение (28)), через некоторые точки на плоскости x, y проходит более одной интегральной линии или не проходит ни одной такой линии. Эти точки называются *особыми точками* для рассматриваемого уравнения. Они могут быть либо *изолированными*, либо заполнять целые *особые линии*.

Начнем с исследования уравнения (36) на простом частном примере

$$y'=y^\alpha \qquad (\alpha>0), \tag{38}$$

причем будем считать $y \geqslant 0$. Уравнение (38) легко интегрируется:

$$\frac{dy}{y^\alpha}=dx, \qquad \frac{y^{1-\alpha}}{1-\alpha}=x-C\ (\alpha\neq 1), \quad \ln y=x-C \quad (\alpha=1); \tag{39}$$

мы написали $-C$ взамен $+C$ для удобства дальнейших рассуждений; это несущественно, так как само C может быть любого знака. Будем различать два случая.

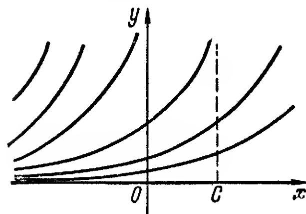


Рис. 288.

1. $\alpha>1$. Тогда решение (39) можно записать в виде

$$y=\frac{1}{(\alpha-1)^{\frac{1}{\alpha-1}}}\cdot\frac{1}{(C-x)^{\frac{1}{\alpha-1}}}=\frac{\text{const}}{(C-x)^{\frac{1}{\alpha-1}}},$$

откуда следует, что если x меняется от $-\infty$ до C, то y возрастает от нуля до бесконечности. С изменением константы C график сдвигается вдоль оси x. Получающееся семейство интегральных линий изображено на рис. 288. Сама ось x также является интегральной линией; она получается в пределе при $C \to \infty$. Как видим, в данном случае через каждую точку верхней полуплоскости проходит единственная интегральная линия. На данном примере видно также, что решение $y(x)$ *может существовать не на всей оси x, а лишь на некоторой ее части* — в данном примере на интервале $-\infty<x<C$.

При $\alpha=1$ получается аналогичная единственность (проверьте!).

2. $0<\alpha<1$. Тогда решение (39) можно записать в виде

$$y=(1-\alpha)^{\frac{1}{1-\alpha}}(x-C)^{\frac{1}{1-\alpha}}=\mathrm{const}\,(x-C)^{\frac{1}{1-\alpha}}, \tag{40}$$

откуда следует, что если x меняется от C до ∞, то y возрастает от нуля до бесконечности. Получающееся семейство интегральных линий изображено на рис. 289. Ось x и здесь является интегральной линией, что видно из уравнения (38); однако она не получается из формулы (40) ни при каком C. Через каждую точку оси x в данном случае проходит по две интегральные линии — сама ось x и кривая, т. е. единственность решения задачи Коши, которая для уравнения первого порядка сводится к задаче о проведении интегральной линии через заданную точку плоскости, в точках оси x нарушается.

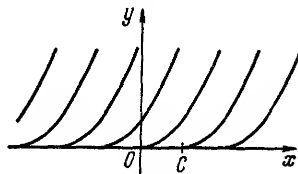


Рис. 289.

Чем «провинились» точки оси x в этом втором случае, что они стали особыми, можно увидеть, вычислив производную от правой части уравнения (38) в этих точках, т. е. при $y=0$

$$\left(\frac{\partial}{\partial y}(y^{\alpha})\right)_{y=0}=(\alpha y^{\alpha-1})_{y=0}=0 \text{ (при } \alpha>1); \ =1 \text{ (при } \alpha=1); \ =\infty \text{ (при } \alpha<1).$$

Поэтому для $0<\alpha<1$ при подходе к оси x поле поворачивается так быстро, что интегральные линии успевают дойти до нее на конечном расстоянии, а не на бесконечности, как на рис. 288.

Итак, мы видим, что в рассматриваемом случае не выполнены условия теоремы Коши (п. 3) о существовании и единственности решения, в которой требовалась конечность производной $\frac{\partial f}{\partial y}$. И в других случаях, если $\left(\frac{\partial f}{\partial y}\right)_{M_0}$ перестает быть конечной, через M_0 может, хоть и не обязательно, пройти более одной интегральной линии.

В частности, если $\frac{\partial f}{\partial y}$ обращается в бесконечность на некоторой линии (L) и сама эта линия является интегральной, то, как правило, через каждую точку (L) проходит, кроме (L), еще по крайней мере одна интегральная линия. В этом случае (L) является *особой интегральной линией*, т. е. интегральной линией, все точки которой особые, а соответствующее решение, графиком которого служит особая интегральная линия, называется *особым решением*. Обычно оно не входит в состав общего решения, т. е. не получается из последнего ни при каком значении произвольной постоянной. Так, в примере (38) при $0<\alpha<1$ ось x служит особой интегральной линией, а функция $y\equiv 0$ — особым решением (почему?).

К особому решению можно подойти с иных позиций. Из рис. 289 видно, что в данном случае ось x служит огибающей (п. XII.5) семейства интегральных линий. Так и в общем случае *огибающая семейства интегральных линий*, если она имеется, *является интегральной линией*, так как она всюду идет вдоль поля, и в то же время *особой*, так как через ее точки проходят и другие интегральные линии. Поэтому если удалось найти общее решение в форме $\Phi(x, y; C)=0$, то на основе п. XII.5 для нахождения особого решения надо исключить C из уравнений

$$\Phi(x, y; C)=0, \quad \Phi'_C(x, y; C)=0. \tag{41}$$

(Проделайте это с решением (40).)

Перейдем теперь к уравнению (37), причем для простоты предположим, что функции P, Q непрерывны, а их частные производные первого порядка конечны. Так как уравнение (37) можно переписать в форме

$$\frac{dy}{dx}=-\frac{P(x, y)}{Q(x, y)} \quad \text{или} \quad \frac{dx}{dy}=-\frac{Q(x, y)}{P(x, y)}, \tag{42}$$

то применима только что упомянутая теорема Коши об уравнении (36), т. е. *через каждую точку* $M_0(x_0, y_0)$, *в которой* $Q(x_0, y_0)\neq 0$ *или* $P(x_0, y_0)\neq 0$, *проходит единственная интегральная линия.* (Для применения теоремы Коши достаточно обозначить через $f(x, y)$ ту из правых частей (42), у которой знаменатель не равен нулю.) Если же

$$P(x_0, y_0)=0, \quad Q(x_0, y_0)=0, \tag{43}$$

то в точке $M_0(x_0; y_0)$ уравнение (37) перестает связывать dx и dy, т. е. поле направлений в такой точке не определено. Поэтому особые точки для уравнения (37) определяются соотношениями (43).

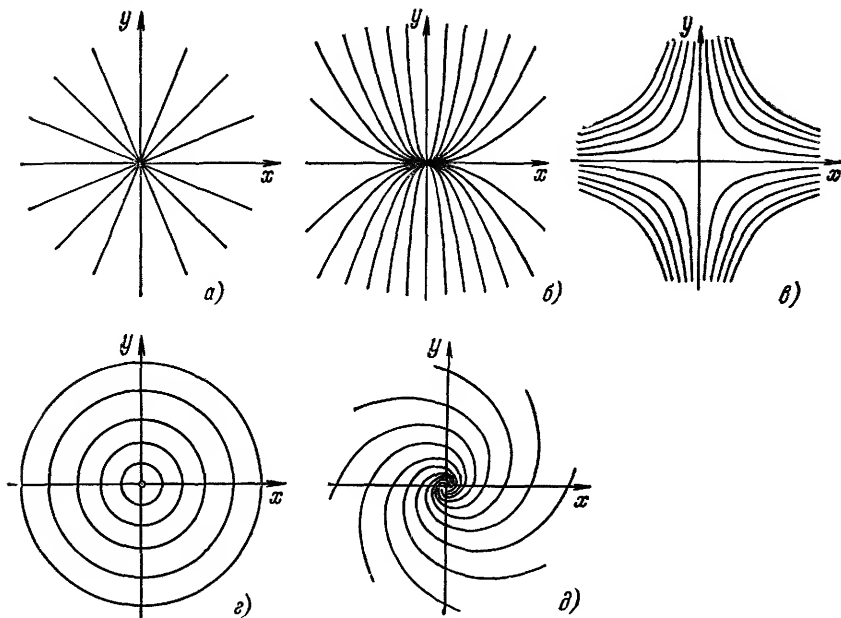


Рис. 290.

Особые точки дифференциальных уравнений:

а) *узел*: $y\,dx - x\,dy=0$, $y=Cx$;
б) *узел*: $2y\,dx - x\,dy=0$, $y=Cx^2$;
в) *седло*: $y\,dx+x\,dy=0$, $xy=C$;
г) *центр*. $x\,dx+y\,dy=0$, $x^2+y^2=C$;
д) *фокус*: $(x+y)\,dx-(x-y)\,dy=0$, $\rho=Ce^{\varphi}$ (в полярных координатах).

Несколько простых наиболее распространенных примеров особых точек вместе с их наименованиями показаны на рис. 290. (Проверьте правильность указанных там общих решений и изображения семейств интегральных линий.)

Во всех этих примерах особой точкой служит начало координат. В примерах *а*, *б* и *д* через особую точку проходит бесконечное количество интегральных линий, в примере *в* — две такие линии и в примере *г* — ни одной. Отметим, что в примерах *а*, *б* и *в* сами оси координат также служат интегральными линиями. При интегрировании уравнения рис. 290, *д* полезно предварительно перейти к полярным координатам.

8. Уравнения, не разрешенные относительно производной. Уравнение

$$F(x,\ y,\ y')=0 \tag{44}$$

отличается от разобранного в п. 3 уравнения (13) тем, что в данном случае y' является неявной функцией x, y. Характерной чертой неявных функций является то, что они, вообще говоря, многозначные (п. I.20). Поэтому если уравнение (44) разрешить относительно y' (в принципе, так как практически это может быть затруднительно), то получится несколько решений:

$$y'=f_1(x,\ y),$$
$$y'=f_2(x,\ y),\ \ldots,\ y'=f_k(x,\ y), \tag{45}$$

любое из которых удовлетворяет уравнению (44).

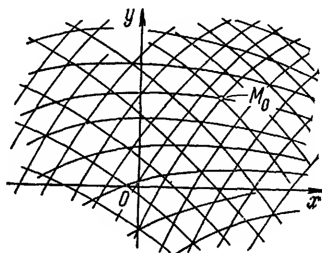


Рис. 291.

Каждое из уравнений (45) задает на плоскости поле направлений и имеет семейство интегральных линий, заполняющее плоскость (п. 3). Поэтому в той части плоскости, где уравнение (44) имеет k решений относительно y', оно определяет k полей направлений, наложенных друг на друга, и через каждую точку проходит k интегральных линий, т. е. начальное условие $y(x_0)=y_0$ определяет k решений (см. рис. 291, где принято $k=3$).

Подобно п. 7, уравнение (44) может иметь особые точки, особые линии и особые интегральные линии. Так как из уравнения (44)

$$\frac{\partial y'}{\partial y}=-\frac{F'_y(x,\ y,\ y')}{F'_{y'}(x,\ y,\ y')}$$

(см. п. IX.13), то в силу теоремы Коши (п. 3) такие точки и линии могут появиться, только если наряду с (44) имеет место равенство

$$F'_{y'}(x,\ y,\ y')=0 \tag{46}$$

(конечно, если $F'_y\neq\infty$).

В частности, особое решение, если оно имеется, график которого является огибающей семейства интегральных линий, можно получить, исключая y' из равенств (44) и (46). Другой способ построения особого решения основан на формулах (41).

9. Метод предварительного дифференцирования. В некоторых случаях уравнение (44) удается проинтегрировать после его *предварительного дифференцирования*. Рассмотрим, например, уравнение

$$x=f(y'), \tag{47}$$

или, как принято записывать,

$$x=f(p)\qquad (p=y'). \tag{48}$$

Если продифференцировать обе части, получим

$$dx = f'(p)\,dp.$$

С помощью этого равенства и формулы $\frac{dy}{dx} = p$ находим выражение для dy:

$$dy = p\,dx = pf'(p)\,dp,$$

откуда

$$y = \int pf'(p)\,dp + C. \tag{49}$$

Равенства (48) и (49) вместе определяют функциональную зависимость между x и y в параметрическом виде (п. II.6), причем параметром служит p. Мы получили общее решение уравнения (47) в параметрическом виде. Аналогично решается уравнение $y = f(y')$ (проверьте!).

Несколько более сложиым является *уравнение Лагранжа*

$$y = f(y')\,x + g(y'), \quad \text{т. е.} \quad y = f(p)\,x + g(p) \qquad (p = y'), \tag{50}$$

линейное относительно x и y, но нелинейное в основном значении этого слова (п. 4). После дифференцирования получаем

$$dy = p\,dx = f'(p)\,dp\,x + f(p)\,dx + g'(p)\,dp,$$

т. е.

$$[p - f(p)]\,\frac{dx}{dp} = f'(p)\,x + g'(p).$$

Если $f(p) \neq p$, то после деления на $p - f(p)$ получается линейное уравнение (см. уравнение (19)), в котором x рассматривается как функция от p. После интегрирования этого уравнения получим равенство вида $x = x(p;\ C)$, которое вместе с (50) даст общее решение исходного уравнения в параметрическом виде.

В частном случае, когда $f(p) \equiv p$, уравнение (50) называется *уравнением Клеро* по имени французского математика А. Клеро́ (1713—1765), впервые рассмотревшего его в 1734 г.; оно имеет вид $y = xy' + g(y')$, т. е.

$$y = xp + g(p) \qquad (p = y'). \tag{51}$$

Предварительное дифференцирование дает $p\,dx = p\,dx + x\,dp + g'(p)\,dp$ т. е.

$$dp\,[x + g'(p)] = 0. \tag{52}$$

Приравнивая нулю первый множитель, получим в силу (51)

$$p = C, \ \text{т. е.} \ y = Cx + g(C). \tag{53}$$

Это — общее решение уравнения (51).

Приравнивая нулю второй множитель в левой части (52), получим

$$x = -g'(p), \quad y = xp + g(p) = -pg'(p) + g(p). \tag{54}$$

Значит, получилось еще одно, особое решение уравнения (51), определенное в параметрическом виде. Геометрически формула (53) задает семейство прямых (почему?), а формулы (54) — огибающую. (Проверьте последнее утверждение, исходя из уравнения (53).)

Например, уравнение $y = xy' - y'^2$ имеет общее решение

$$y = Cx - C^2 \tag{55}$$

и особое решение, графиком которого служит огибающая семейства прямых (55). Для ее отыскания продифференцируем по C обе части (55), что даст

$$0 = x - 2C.$$

Исключая C из двух последних формул, получим $C = \frac{x}{2}$, т. е.

$$y = \frac{x}{2}x - \left(\frac{x}{2}\right)^2 = \frac{x^2}{4}.$$

Соответствующие интегральные линии показаны на рис. 292.

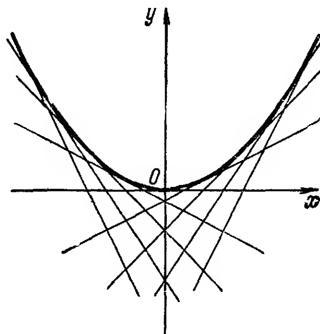


Рис. 292

§ 3. Уравнения высших порядков и системы уравнений

10. Уравнения высших порядков. Общие понятия, относящиеся к таким уравнениям, были приведены в п. 2 (уравнение (5), общее решение (8) или (9), начальное условие (11)). Впрочем, как и для первого порядка, уравнение порядка n обычно бывает проще исследовать, если оно задано в форме, разрешенной относительно старшей производной:

$$y^{(n)} = f(x, y, y', \ldots, y^{(n-1)}).$$

В частности, на эту форму непосредственно распространяется теорема Коши (п. 3): *начальные значения* (11) *определяют одно и только одно решение, если при этих значениях функция f непрерывна и имеет конечные производные первого порядка по* $y, y', \ldots, y^{(n-1)}$.

Рассмотрим вопрос об интегрировании этих уравнений в квадратурах; в случае уравнений высшего порядка интегрирование удается довести до конца еще реже, чем для уравнений первого порядка. Основным способом формального интегрирования нелинейных уравнений высшего порядка (о линейных уравнениях мы будем говорить особо в § 4) является *метод понижения порядка*, т. е. переход к равносильному уравнению низшего порядка. Как правило, чем ниже порядок уравнения, тем оно проще. Кроме того, бывает, что после одного или нескольких понижений порядка мы переходим к уравнению первого порядка одного из интегрируемых типов (п. 4); тогда интегрирование удается довести до конца. Рассмотрим некоторые частные способы понижения порядка; некоторые другие способы указаны в [22].

1. Пусть, например, задано уравнение второго порядка

$$y'^2 + yy'' = 0.$$

Для его интегрирования заметим, что левую часть можно переписать в виде

$$y'^2 + yy'' \equiv (yy')',$$

откуда

$$(yy')' = 0;\ \ yy' = C_1;\ \ y\,dy = C_1 dx;\ \ \frac{y^2}{2} = C_1 x + C_2 \quad \text{(общее решение)}.$$

Другое уравнение

$$y'^2 - yy'' = 0$$

легко проинтегрировать аналогичным образом, если предварительно разделить обе части на y^2:

$$\frac{y'^2 - yy''}{y^2} = 0;\ \ -\left(\frac{y'}{y}\right)' = 0;\ \ \frac{y'}{y} = C_1;\ \ \frac{dy}{y} = C_1 dx;$$

$$\ln|y| = C_1 x + \ln C_2;\quad y = C_2 e^{C_1 x} \quad \text{(общее решение)}.$$

Как говорят, после деления на y^2 мы получили *интегрируемую комбинацию*: *«точная производная» приравнена к нулю*. Аналогичный прием иногда применяется и в других примерах.

Дальнейшие случаи понижения порядка мы будем для простоты излагать для уравнений второго порядка общего вида

$$F(x,\ y,\ y',\ y'') = 0. \tag{56}$$

2. *Пусть в уравнении* (56) *не присутствует* y, а только производные от него, т. е. мы имеем уравнение вида

$$F(x,\ y',\ y'') = 0. \tag{57}$$

Тогда вводят обозначение $y' = p = p(x)$ и из (57) получится

$$F(x,\ p,\ p') = 0,$$

т. е. уравнение первого порядка. Если нам повезет и удастся его проинтегрировать, то получим общее решение уравнения (57)

$$p = \varphi(x;\ C_1);\quad y' = \varphi(x;\ C_1);\quad y = \int \varphi(x;\ C_1)\,dx + C_2.$$

3. *Пусть в уравнении* (56) *не присутствует* x, т. е. мы имеем уравнение

$$F(y,\ y',\ y'') = 0. \tag{58}$$

Тогда также обозначают $y' = p$, но рассматривают p как функцию от y. При этом в (58) нельзя подставлять просто $y'' = p'$, так как тогда p' означало бы производную от p по x, а не по y. Поэтому пишут

$$y'' = \frac{d(y')}{dx} = \frac{dp}{dx} = \frac{dp}{dy}\cdot\frac{dy}{dx} = p\frac{dp}{dy}.$$

Из уравнения (58) получим

$$F\left(y,\ p,\ p\frac{dp}{dy}\right) = 0,$$

т. е. уравнение первого порядка. Если его удастся проинтегрировать, то мы сможем найти общее решение уравнения (58)

$$p=\varphi(y;\ C_1);\quad \frac{dy}{dx}=\varphi(y;\ C_1);\quad \int\frac{dy}{\varphi(y;\ C_1)}=x+C_2.$$

4. *Пусть левая часть уравнения* (56) *однородна относительно неизвестной функции и ее производных*, т. е. (см. п. IX.12)

$$F(x,\ ty,\ ty',\ ty'')\equiv t^kF(x,\ y,\ y',\ y''). \tag{59}$$

В этом случае порядок понижается после подстановки

$$\frac{y'}{y}=u=u(x),$$

откуда

$$y'=uy,\quad y''=u'y+uy'=u'y+u\cdot uy=(u'+u^2)\,y,$$
$$F(x,\ y,\ yu,\ y(u'+u^2))=0,\quad F(x,\ 1,\ u,\ u'+u^2)=0;$$

в последнем переходе мы воспользовались свойством (59). Если полученное уравнение первого порядка удастся проинтегрировать, то

$$u=\varphi(x;\ C_1);\quad \frac{y'}{y}=\varphi(x;\ C_1);$$
$$\ln|y|=\int\varphi(x;\ C_1)\,dx+\ln C_2;\quad y=C_2e^{\int\varphi(x;\ C_1)\,dx}.$$

11. Связь уравнений высшего порядка с системами уравнений первого порядка. От уравнения высшего порядка (5) всегда можно перейти к равносильной системе n уравнений первого порядка с n неизвестными функциями. Для этого надо обозначить

$$y=y_1,\ y'=y_2,\ \ldots,\ y^{(n-1)}=y_n. \tag{60}$$

Тогда в силу уравнения (5) можно написать

$$\left.\begin{aligned} &y_1'=y_2,\\ &y_2'=y_3,\\ &\ldots\ldots\ldots\ldots\ldots\\ &y_{n-1}'=y_n,\\ &F(x,\ y_1,\ y_2,\ \ldots,\ y_n,\ y_n')=0.\end{aligned}\right\} \tag{61}$$

Система (61) имеет специальный вид. Общий вид системы первого порядка, например, из трех уравнений с тремя неизвестными функциями таков:

$$\left.\begin{aligned} &F_1(x,\ y_1,\ y_2,\ y_3,\ y_1',\ y_2',\ y_3')=0,\\ &F_2(x,\ y_1,\ y_2,\ y_3,\ y_1',\ y_2',\ y_3')=0,\\ &F_3(x,\ y_1,\ y_2,\ y_3,\ y_1',\ y_2',\ y_3')=0.\end{aligned}\right\} \tag{62}$$

Обратно, из системы (62) можно перейти к одному уравнению (в данном случае третьего порядка) с одной неизвестной функцией, например y_1. Поэтому общее решение системы (62) содержит три произвольные постоянные:

$$y_1=\varphi_1(x;\ C_1,\ C_2,\ C_3),\ y_2=\varphi_2(x;\ C_1,\ C_2,\ C_3),$$
$$y_3=\varphi_3(x;\ C_1,\ C_2,\ C_3).$$

Для указанного перехода надо каждое из уравнений (62) продифференцировать два раза и из полученных девяти равенств, считая вместе с равенствами (62), исключить восемь величин y_2, y_3, y_2', y_3', y_2'', y_3'', y_2''', y_3'''. После интегрирования получается общее решение $y_1 = \varphi_1(x;\ C_1,\ C_2,\ C_3)$, из которого y_2 и y_3 находятся уже без интегрирований на основе упомянутых равенств.

Систему уравнений любого порядка можно преобразовать в систему уравнений первого порядка с помощью обозначений, аналогичных (60). Например, система двух уравнений, третьего порядка относительно одной неизвестной функции и второго порядка относительно другой, эквивалентна системе из пяти уравнений первого порядка с пятью неизвестными функциями; общее решение в этом примере содержит пять произвольных постоянных.

12. Геометрический смысл системы уравнений первого порядка. Рассмотрим для простоты случай системы из двух уравнений первого порядка с двумя неизвестными функциями $y_1(x)$ и $y_2(x)$:

$$\left.\begin{aligned} F_1(x,\ y_1,\ y_2,\ y_1',\ y_2') = 0, \\ F_2(x,\ y_1,\ y_2,\ y_1',\ y_2') = 0. \end{aligned}\right\} \tag{63}$$

Если эту систему до интегрирования удастся разрешить относительно y_1' и y_2', то она примет более простой вид

$$\left.\begin{aligned} y_1' = f_1(x,\ y_1,\ y_2), \\ y_2' = f_2(x,\ y_1,\ y_2); \end{aligned}\right\} \tag{64}$$

тогда говорят, что система записана в *нормальной форме*.

Решением системы (63) или, что то же, (64) называется, конечно, *пара* функций

$$y_1 = y_1(x), \quad y_2 = y_2(x), \tag{65}$$

обращающая *оба* уравнения в тождества. В соответствии с п. 11 в *общее решение* входят две произвольные постоянные, т. е. оно имеет вид

$$y_1 = y_1(x;\ C_1,\ C_2), \quad y_2 = y_2(x;\ C_1,\ C_2).$$

Система уравнений (64) и ее решения (65) имеют простой геометрический смысл, для выяснения которого надо рассмотреть трехмерное пространство x, y_1, y_2. Тогда формулы (65) определят линию в параметрическом виде (п. VII. 23), причем здесь параметром служит сам x (можно дописать равенство $x = x$); она называется *интегральной линией* системы уравнений (64). Если для произвольной точки M в пространстве (рис. 293) подсчитать значения правых частей системы (64), то мы будем знать направления касательных к линиям $y_1 = y_1(x)$ и $y_2 = y_2(x)$, т. е. к проекциям интегральной линии, и тем самым сможем узнать направление касательной к самой интегральной линии, если она проходит через M. Значит, система (64) задает *поле направлений в пространстве* x, y_1, y_2, а интегральная линия — это линия, в каждой своей точке идущая «вдоль поля», т. е. линия, в каждой точке которой касательная имеет направление, заданное этим полем (ср. с п. 3).

В системе (64) переменные y_1 и y_2 равноправны, а переменная x имеет иное значение. Бывает, что все три переменные равноправны, так что любую

из них можно принять за независимую; тогда систему уравнений предпочитают записывать в *симметричной форме*, например,

$$\frac{dx}{P(x,\ y,\ z)}=\frac{dy}{Q(x,\ y,\ z)}=\frac{dz}{R(x,\ y,\ z)}. \tag{66}$$

От формы (66) легко перейти к форме (64) и наоборот (как?).

Геометрический смысл системы (66) аналогичен описанному выше. Так как вектор $d\mathbf{r}=dx\mathbf{i}+dy\mathbf{j}+dz\mathbf{k}$ в любой заданной точке $M(x;\ y;\ z)$ в силу соотношений (66) должен быть параллелен известному вектору $P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$ (почему?), то задача об интегрировании системы (66) — это задача о построении линий в пространстве, имеющих в каждой своей точке заданное направление.

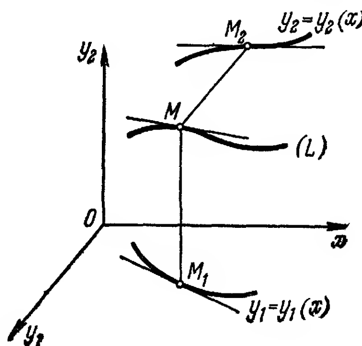


Рис. 293.
(L) — интегральная линия.

Из геометрического смысла системы (64) вытекает, что для однозначного определения интегральной линии надо задать точку $M_0(x_0;\ y_{10};\ y_{20})$ в пространстве, через которую эта линия должна пройти. Другими словами, *начальное условие*

$$y_1(x_0)=y_{10},\quad y_2(x_0)=y_{20}$$

однозначно определяет решение системы (64). Конечно, как и в п. 7, и здесь возможны особые точки и особые линии, которые распознаются в общем подобно п. 7. В частности, *для системы* (66) *особой точкой является всякая точка, в которой все три знаменателя обращаются в нуль*, т. е. вектор $P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$ обращается в нуль-вектор, не имеющий определенного направления.

Система первого порядка в нормальной форме с любым числом уравнений имеет общий вид

$$\left.\begin{aligned} y_1' &= f_1(x,\ y_1,\ y_2,\ \ldots,\ y_n),\\ y_2' &= f_2(x,\ y_1,\ y_2,\ \ldots,\ y_n),\\ &\cdots\cdots\cdots\cdots\\ y_n' &= f_n(x,\ y_1,\ y_2,\ \ldots,\ y_n). \end{aligned}\right\} \tag{67}$$

Решение ее — это система функций

$$y_1=y_1(x),\quad y_2=y_2(x),\ \ldots,\ y_n=y_n(x). \tag{68}$$

Общее решение содержит n произвольных постоянных.

Для однозначного определения частного решения можно задать начальное условие

$$y_1(x_0)=y_{10};\quad y_2(x_0)=y_{20},\ \ldots,\ y_n(x_0)=y_{n0}. \tag{69}$$

Коши доказал, что условиям (69) удовлетворяет ровно одно решение системы (67), если при значениях $x=x_0,\ y_1=y_{10},\ \ldots,\ y_n=y_{n0}$ правые части системы (67) непрерывны, а их производные по переменным $y_1,\ y_2,\ \ldots,\ y_n$ конечны.

Геометрический смысл системы (67), решения (68) и условий (69) — это соответственно поле направлений, интегральная линия и точка, через которую должна пройти эта линия в $(n+1)$-мерном пространстве $x, y_1, y_2, \ldots, y_n$ (п VII.18).

Если правые части системы (67) не содержат независимой переменной x, то эта система называется *автономной*; оказывается, что ее решения удобнее рассматривать в n-мерном пространстве $y_1, y_2, \ldots, y_n$, называемом *фазовым пространством*. Мы ограничимся для простоты случаем $n=2$, будем обозначать независимую переменную буквой t и *истолковывать ее как время*, а искомые функции взамен y_1, y_2 будем обозначать x, y, так что $x=x(t)$, $y=y(t)$. Вместо (67) тогда получится система уравнений

$$\frac{dx}{dt}=P(x, y), \quad \frac{dy}{dt}=Q(x, y).$$

Если умножить первое уравнение на $\mathbf{i}$, второе — на $\mathbf{j}$, а затем произвести почленное сложение, мы получим векторное дифференциальное уравнение

$$\frac{d\mathbf{r}}{dt}=\mathbf{A}(x, y) \quad (=\mathbf{A}(\mathbf{r})), \tag{70}$$

где $\mathbf{A}=P(x, y)\mathbf{i}+Q(x, y)\mathbf{j}$ — заданное векторное поле на *фазовой плоскости* x, y. Так как $\frac{d\mathbf{r}}{dt}$ — это вектор скорости (п. VII.23), то на плоскости x, y оказывается заданным *поле скоростей*, а решение $\mathbf{r}(t)=x(t)\mathbf{i}+y(t)\mathbf{j}$ определяет закон движения точки на плоскости, при котором эта точка в каждом своем положении имеет скорость, заданную для этого положения. Несколько вольно можно представлять себе, что уравнение (70) задает на фазовой плоскости поток жидкости, а решениям отвечают законы движения частиц этой жидкости. Автономность уравнения (70) означает, что рассматриваемый поток стационарный, а потому различные траектории не имеют друг с другом общих точек.

Запишем, например, уравнение (4) в виде автономной системы первого порядка

$$\frac{dy}{dt}=v, \quad M\frac{dv}{dt}=-ky; \tag{71}$$

здесь y и v — координата и скорость колеблющейся точки. В курсе физики выводится выражение для полной энергии колеблющейся точки

$$E=\frac{Mv^2}{2}+\frac{ky^2}{2} \tag{72}$$

(продумайте эту формулу). При свободных колебаниях без трения энергия должна сохраняться. И действительно, в силу (71)

$$\frac{dE}{dt}=Mv\frac{dv}{dt}+ky\frac{dy}{dt}=-kyv+kyv=0;$$

это — математическое доказательство *закона сохранения энергии* в данном примере. Таким образом, $E=\text{const}$ для любого решения системы (71), т. е. движения в фазовой плоскости y, v происходят по эллипсам, причем разным эллипсам отвечают колебания вокруг положения равновесия с различной амплитудой (рис. 294).

13. Первые интегралы. Рассмотрим для определенности систему из трех уравнений первого порядка вида (62). Всякое соотношение вида

$$\Phi(x, y_1, y_2, y_3; C)=0, \tag{73}$$

обязанное тождественно удовлетворяться для любого решения системы, называется *первым интегралом* этой системы уравнений; здесь C — постоянная, вообще говоря, различная для различных решений. *Знание первого интеграла дает возможность понизить число уравнений в системе на единицу*: например, если из (73) выразить y_3 через все остальное и подставить результат в первые два уравнения (62), то получится система из двух уравнений первого порядка с двумя неизвестными функциями y_1 и y_2. Если ее проинтегрировать, т. е. найти $y_1(x)$ и $y_2(x)$, то $y_3(x)$ можно будет найти без интегрирований из равенства (73).

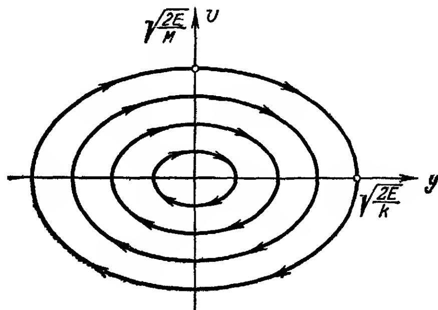


Рис. 294.

Аналогичным образом знание двух независимых первых интегралов позволяет понизить число уравнений на два, а три независимых первых интеграла (т. е. таких, что ни один из них не является следствием остальных)

$$\left.\begin{aligned} &\Phi_1(x, y_1, y_2, y_3; C_1)=0, \\ &\Phi_2(x, y_1, y_2, y_3; C_2)=0, \\ &\Phi_3(x, y_1, y_2, y_3; C_3)=0 \end{aligned}\right\}$$

дают общее решение системы (62), записанное в неявной форме.

Иногда первые интегралы удается найти, выводя из заданных уравнений системы *интегрируемые комбинации*. Например, для системы

$$\left.\begin{aligned} y' &= y+z, \\ z' &= -y+z \end{aligned}\right\} \tag{74}$$

легко получить такую комбинацию:

$$yy'+zz' = y(y+z)+z(-y+z) = y^2+z^2,$$

т. е.

$$\frac{1}{2}(y^2+z^2)' = y^2+z^2, \quad \frac{d(y^2+z^2)}{y^2+z^2} = 2dx, \quad \ln(y^2+z^2) = 2x+\ln C,$$

и окончательно имеем первый интеграл

$$y^2+z^2 = Ce^{2x}.$$

Из него видно, например, что при $x \to \infty$ решение уходит в бесконечность, а при $x \to -\infty$ решение стремится к нулю; и в других случаях бывает возможно сделать существенные выводы о поведении решений без полного интегрирования системы. Еще один первый интеграл для системы (74) можно получить, разделив одно из уравнений (74) на другое (проделайте это!).

В некоторых случаях первые интегралы подсказываются физическими соображениями, чаще всего теми или иными законами сохранения.

Например, формула (72), в которой E играет роль произвольной постоянной C, служит первым интегралом системы (71). Выразив из него v через y и подставив результат в первое уравнение (71), легко довести интегрирование до конца (проделайте это!).

Подчеркнем в заключение, что, как видно из предыдущего, наиболее естественно рассматривать системы, в которых число уравнений равно числу неизвестных функций; такие системы принято называть *замкнутыми*. Если

уравнений меньше, чем искомых функций, то система называется *незамкнутой* (*недоопределенной*); у такой системы избыточное количество неизвестных функций можно задавать произвольно. Чаще всего незамкнутость системы свидетельствует о том, что просто не все необходимые соотношения выписаны. Если уравнений больше, чем неизвестных функций, то система называется *переопределенной*; такая система обычно противоречива, т. е. не имеет решений. Переопределенность системы обычно свидетельствует либо о ее зависимости, т. е. о том, что некоторые из уравнений являются следствиями остальных и потому излишни, либо об ошибке при ее составлении.

§ 4. Линейные уравнения общего вида

14. Линейные однородные уравнения. Исследование линейных уравнений любого порядка во многом аналогично исследованию линейных уравнений первого порядка (п. 4), хотя теперь уже получить решение в квадратурах в общем случае не удастся. Рассмотрим сначала для простоты уравнение второго порядка. Уравнение

$$z''+p(x)z'+q(x)z=0, \qquad (75)$$

левая часть которого линейна относительно неизвестной функции и ее производных, называется *линейным однородным уравнением*.

Обозначим для краткости левую часть уравнения (75) через $\mathrm{L}[z]$, т. е. в данном случае

$$\mathrm{L}[z]\equiv z''+p(x)z'+q(x)z \text{ (по определению).}$$

Тогда уравнение (75) можно переписать в виде

$$\mathrm{L}[z]=0.$$

Выражение $\mathrm{L}[z]$ обладает следующими свойствами:

$$\mathrm{L}[z_1+z_2]=(z_1+z_2)''+p(x)(z_1+z_2)'+q(x)(z_1+z_2)=$$
$$=(z_1''+p(x)z_1'+q(x)z_1)+(z_2''+p(x)z_2'+q(x)z_2)=\mathrm{L}[z_1]+\mathrm{L}[z_2],$$

$$\mathrm{L}[Cz]=C\mathrm{L}[z] \quad (C=\text{const}) \text{ (проверяется аналогично).}$$

О таких выражениях, называемых линейными операторами, мы упоминали в п. XIV. 26.

Легко доказать следующие свойства уравнения (75).

1. *Сумма решений уравнения* (75) *будет решением того же уравнения.* Действительно, если z_1 и z_2 — два таких решения, т. е.

$$\mathrm{L}[z_1]=0 \text{ и } \mathrm{L}[z_2]=0, \text{ то } \mathrm{L}[z_1+z_2]=\mathrm{L}[z_1]+\mathrm{L}[z_2]=0.$$

Аналогично проверяется свойство 2:

2. *Если решение уравнения* (75) *умножить на константу, то получится решение того же уравнения.*

Свойства 1 и 2 можно объединить так: линейная комбинация (п. VII.5) решений уравнения (75) будет решением того же уравнения. Например, если $z_1(x)$ и $z_2(x)$ удовлетворяют уравнению (75),

то и

$$z = C_1 z_1(x) + C_2 z_2(x) \tag{76}$$

удовлетворяет тому же уравнению при любых постоянных C_1, C_2.

3. *Тождественно нулевая функция удовлетворяет уравнению* (75).

4. *Если известно ненулевое решение уравнения* (75), *то его порядок можно понизить на единицу.* Действительно, пусть $z_1(x)$ — такое решение; сделаем подстановку $z = z_1 u$, где $u = u(x)$ — новая неизвестная функция. Получим

$$(z_1'' u + 2z_1' u' + z_1 u'') + p(z_1' u + z_1 u') + q z_1 u = 0,$$

т. е.

$$z_1 u'' + (2z_1' + p z_1) u' + (z_1'' + p z_1' + q z_1) u = 0.$$

Но так как $L[z_1] = 0$, то последний член отпадает и после подстановки $u' = v$ получаем

$$z_1 v' + (2z_1' + p z_1) v = 0,$$

т. е. линейное однородное уравнение на единицу низшего, чем было, порядка.

Доведем интегрирование до конца:

$$\frac{dv}{v} = -\frac{2z_1' + p z_1}{z_1} dx, \quad \ln|v| = -2\ln|z_1| - \int p(x)\,dx + \ln C_2,$$

$$v = \frac{C_2}{z_1^2} e^{-\int p(x)\,dx}, \quad u = C_2 \int \frac{1}{z_1^2} e^{-\int p(x)\,dx} dx + C_1,$$

$$z = C_1 z_1 + C_2 z_1 \int \frac{1}{z_1^2} e^{-\int p(x)\,dx} dx. \tag{77}$$

Функция, при которой стоит множитель C_2, является одним из частных решений уравнения (75), так как она получается из общего решения (77), если положить $C_1 = 0$, $C_2 = 1$. Поэтому если обозначить ее через z_2, то мы приходим к свойству 5:

5. *Общее решение уравнения* (75) *имеет вид* (76), где C_1 и C_2 — произвольные постоянные, а z_1 и z_2 — два частных решения этого уравнения.

В этом свойстве *в качестве* z_1, z_2 *могут быть взяты только два линейно независимых решения*, а не любая пара решений. Понятие линейной зависимости функций вводится подобно аналогичному понятию для векторов (п. VII.5): именно, несколько функций называются *линейно зависимыми* друг от друга, если одна из них является линейной комбинацией остальных. В частности, две функции $z_1(x)$ и $z_2(x)$ линейно зависимы, если $z_2(x) \equiv C z_1(x)$, т. е. если они пропорциональны. Тогда формула (76) не дает общего решения, так как

$$C_1 z_1 + C_2 z_2 \equiv C_1 z_1 + C_2 C z_1 \equiv (C_1 + C_2 C) z_1(x) = D z_1(x),$$

где $D = C_1 + C_2 C$ — постоянная; значит, хотя формально справа имеются две произвольные постоянные, но они не являются существенными, т. е. их число можно уменьшить на единицу (ср. п. X.2).

Все указанные свойства справедливы и для линейного однородного уравнения любого порядка

$$z^{(n)}+p(x)z^{(n-1)}+q(x)z^{(n-2)}+\ldots+s(x)z=0, \qquad (78)$$

за тем исключением, что общее решение, взамен (76), имеет вид

$$z=C_1z_1(x)+C_2z_2(x)+\ldots+C_nz_n(x). \qquad (79)$$

Здесь все C — произвольные постоянные, а $z_1, z_2, \ldots, z_n$ — какие-либо линейно независимые решения уравнения (78). Совокупность n линейно независимых решений уравнения (78) порядка n называется *фундаментальной системой решений*. Таким образом, *общее решение уравнения* (78) *есть линейная комбинация решений из фундаментальной системы с произвольными коэффициентами*. В терминах пп. VII.17—19 можно сказать, что *совокупность всех решений уравнения* (78) *образует n-мерное линейное пространство*; *фундаментальная система решений — это базис в этом пространстве.*

Отметим в заключение, что у уравнения (78) можно понизить порядок на единицу по методу п. 10 (тип 4), но это делают редко, так как после понижения порядка уравнение становится нелинейным.

15. Неоднородные уравнения. Рассмотрим теперь *линейное неоднородное уравнение*

$$y''+p(x)y'+q(x)y=f(x). \qquad (80)$$

В соответствии с п. 14 обозначим левую часть через $\mathrm{L}[y]$.

1. *Знание какого-либо частного решения уравнения* (80) *позволяет свести задачу об интегрировании этого уравнения к задаче об интегрировании соответствующего* (т. е. с отброшенной правой частью) *однородного уравнения* (75).

Действительно, если $Y(x)$ — такое решение, то, сделав замену

$$y=Y(x)+z, \qquad (81)$$

где $z=z(x)$ — новая неизвестная функция, получим

$$\mathrm{L}[Y+z]=f(x), \quad \mathrm{L}[Y]+\mathrm{L}[z]=f(x).$$

Однако $\mathrm{L}[Y]=f(x)$ (почему?), и мы получаем уравнение (75) для z.

Итак, *общее решение линейного неоднородного уравнения* (80) *есть сумма какого-либо его частного решения и общего решения соответствующего однородного уравнения* (ср. с решением линейного уравнения в п. 4).

2. *Если правая часть $f(x)$ равна линейной комбинации, например, двух функций, т. е. $f(x)=\alpha f_1(x)+\beta f_2(x)$* ($\alpha, \beta=\mathrm{const}$), *и известны какие-либо частные решения $Y_1(x)$ и $Y_2(x)$ уравнения* (80) *с правыми частями $f_1(x)$ и $f_2(x)$, то функция*

$$Y(x)=\alpha Y_1(x)+\beta Y_2(x)$$

служит частным решением уравнения (80) *с правой частью $f(x)$.*

Доказательство этого простого факта, который служит частным проявлением *принципа суперпозиции* (см. п. XIV.26), мы предоставим читателю.

3. *Если известно общее решение однородного уравнения* (75), *то общее решение уравнения* (80) *можно найти с помощью квадратур*.

Это делается с помощью найденного Лагранжем *метода вариации произвольных постоянных* следующим образом. Как мы знаем, общее решение уравнения (75) имеет вид (76). По аналогии с формулой (22) мы ищем решение уравнения (80) в виде

$$y=\varphi_1(x)\,z_1(x)+\varphi_2(x)\,z_2(x), \tag{82}$$

где φ_1, φ_2 — некоторые неизвестные пока функции. Так как их две, а уравнение (80) одно, то для нахождения этих функций мы наложим на них еще одно дополнительное соотношение ((84)).

Дифференцируя равенство (82), получим

$$y'=(\varphi_1 z_1'+\varphi_2 z_2')+(\varphi_1' z_1+\varphi_2' z_2). \tag{83}$$

Потребуем, чтобы вторая скобка обратилась в нуль:

$$\varphi_1' z_1+\varphi_2' z_2=0. \tag{84}$$

Тогда при дифференцировании равенства (83) надо принимать во внимание только первую скобку, т. е.

$$y''=(\varphi_1 z_1''+\varphi_2 z_2'')+(\varphi_1' z_1'+\varphi_2' z_2'). \tag{85}$$

Подставляем все полученные результаты (82), (83) и (85) в уравнение (80), конечно, не выписывая нулевой суммы. Это даст (проверьте!)

$$\varphi_1(z_1''+pz_1'+qz_1)+\varphi_2(z_2''+pz_2'+qz_2)+(\varphi_1' z_1'+\varphi_2' z_2')=f(x).$$

Поскольку функции z_1, z_2 удовлетворяют уравнению (75), то в последнем уравнении первые две скобки отпадают и оно превращается в равенство

$$\varphi_1' z_1'+\varphi_2' z_2'=f(x). \tag{86}$$

Итак, для нахождения φ_1, φ_2 у нас остались два соотношения: (84) и (86). Так как z_1, z_2 и $f(x)$ считаются известными, то получается система двух алгебраических уравнений первой степени с двумя неизвестными: φ_1', φ_2'. Решая систему, мы находим эти неизвестные, а интегрируя, находим φ_1, φ_2.

Рассмотрим, например, простейшее *уравнение вынужденных колебаний*, которое получится, если в правой части уравнения (3) добавить внешнюю силу $P(t)$. Разделив обе части уравнения на M, получим

$$y''+\omega_0^2 y=f(t), \tag{87}$$

где обозначено $\omega_0^2=\frac{k}{M}$, $f(t)=\frac{P(t)}{M}$.

Соответствующее однородное уравнение

$$z'' + \omega_0^2 z = 0 \tag{88}$$

имеет, как легко непосредственно проверить, два решения, $z_1 = \cos\omega_0 t$, $z_2 = \sin\omega_0 t$, и тем самым общее решение

$$z = C_1 \cos\omega_0 t + C_2 \sin\omega_0 t. \tag{89}$$

Отсюда видно, в частности, что ω_0 — это *собственная частота* колебаний рассматриваемой системы, т. е. частота колебаний при отсутствии внешних сил.

Согласно формуле (82) решение уравнения (87) ищем в виде

$$y = \varphi_1(t)\cos\omega_0 t + \varphi_2(t)\sin\omega_0 t. \tag{90}$$

Тогда уравнения (84) и (86) приобретают вид

$$\left.\begin{aligned} &\varphi_1' \cos\omega_0 t + \varphi_2' \sin\omega_0 t = 0,\\ &\varphi_1'(-\omega_0 \sin\omega_0 t) + \varphi_2'\omega_0 \cos\omega_0 t = f(t). \end{aligned}\right\}$$

Отсюда непосредственно находим

$$\varphi_1'(t) = -\frac{1}{\omega_0} f(t)\sin\omega_0 t, \quad \varphi_2'(t) = \frac{1}{\omega_0} f(t)\cos\omega_0 t.$$

При интегрировании здесь неудобно воспользоваться неопределенным интегралом из-за наличия в нем неуточняемой произвольной постоянной; лучше нижний предел интеграла зафиксировать, например, положив его равным моменту $t = 0$ начала отсчета времени:

$$\varphi_1(t) = -\frac{1}{\omega_0}\int_0^t f(t)\sin\omega_0 t\, dt + C_1,$$

где C_1 — произвольная постоянная. Поскольку в правой части t имеет два смысла — переменная интегрирования и верхний предел, то лучше воспользоваться независимостью определенного интеграла от обозначения переменной интегрирования (п. XIV.3) и написать

$$\varphi_1(t) = -\frac{1}{\omega_0}\int_0^t f(\tau)\sin\omega_0\tau\, d\tau + C_1;$$

$$\varphi_2(t) = \frac{1}{\omega_0}\int_0^t f(\tau)\cos\omega_0\tau\, d\tau + C_2.$$

Подставляя в (90), получим

$$y = -\frac{1}{\omega_0}\cos\omega_0 t\int_0^t f(\tau)\sin\omega_0\tau\, d\tau + \\ + \frac{1}{\omega_0}\sin\omega_0 t\int_0^t f(\tau)\cos\omega_0\tau\, d\tau + C_1\cos\omega_0 t + C_2\sin\omega_0 t.$$

Теперь внесем $\cos\omega_0 t$ и $\sin\omega_0 t$ под знак интеграла (этого нельзя было бы сделать, не переменив обозначения переменной интегрирования t на τ) и объединим оба интеграла:

$$y=\frac{1}{\omega_0}\int_0^t [-f(\tau)\cos\omega_0 t\sin\omega_0\tau+f(\tau)\sin\omega_0 t\cos\omega_0\tau]\,d\tau+$$
$$+C_1\cos\omega_0 t+C_2\sin\omega_0 t.$$

Отсюда получаем общее решение уравнения (87):

$$y=\frac{1}{\omega_0}\int_0^t \sin\omega_0(t-\tau)f(\tau)\,d\tau+C_1\cos\omega_0 t+C_2\sin\omega_0 t. \qquad (91)$$

Произвольные постоянные C_1 и C_2 можно определить, например, из начальных условий

$$y(0)=y_0, \quad y'(0)=v_0. \qquad (92)$$

Подставляя в обе части (91) значение $t=0$, получим $y_0=C_1$. Чтобы использовать второе условие (92), надо продифференцировать равенство (91) по t, после чего подставить $t=0$. При дифференцировании интеграла надо иметь в виду, что t входит в него дважды, т. е. (см. формулу (XIV.80))

$$y'=\frac{1}{\omega_0}\int_0^t \omega_0\cos\omega_0(t-\tau)f(\tau)\,d\tau+$$
$$+\left[\frac{1}{\omega_0}\sin\omega_0(t-\tau)f(\tau)\right]_{\tau=t}-C_1\omega_0\sin\omega_0 t+C_2\omega_0\cos\omega_0 t=$$
$$=\int_0^t \cos\omega_0(t-\tau)f(\tau)\,d\tau-C_1\omega_0\sin\omega_0 t+C_2\omega_0\cos\omega_0 t;$$
$$v_0=C_2\omega_0; \quad C_2=\frac{v_0}{\omega_0}.$$

Отсюда решение уравнения (87) при начальных условиях (92)

$$y=\frac{1}{\omega_0}\int_0^t \sin\omega_0(t-\tau)f(\tau)\,d\tau+y_0\cos\omega_0 t+\frac{v_0}{\omega_0}\sin\omega_0 t.$$

Все указанные свойства справедливы и для уравнения

$$y^{(n)}+p(x)\,y^{(n-1)}+q(x)\,y^{(n-2)}+\ldots+s(x)\,y=f(x). \qquad (93)$$

Метод вариации произвольных постоянных будет выглядеть так: в формулу (79) вместо $C_1, C_2, \ldots, C_n$ надо подставить $\varphi_1(x), \varphi_2(x), \ldots \ldots, \varphi_n(x)$, после чего последовательно дифференцировать эту формулу, приравнивая на каждом шаге, вплоть до $(n-1)$-го, получающуюся группу членов с φ'_k нулю; n-е соотношение получится из подстановки всех полученных выражений в (93).

4. *Любое решение уравнения* (78), *а также* (93) *можно продолжить на любой интервал, на котором коэффициенты и правая часть не обращаются в бесконечность.* Для нелинейных уравнений может получиться, что решение или его производные при таком продолжении уходят в бесконечность для конечного значения x. Простым примером этого служит уравнение (38) при $\alpha > 1$ (рис. 288): здесь с ростом y поле направлений так быстро увеличивает крутизну, что интегральная линия поднимается в бесконечность, пройдя лишь конечное расстояние вдоль оси x. В отличие от этого для линейного уравнения, например, вида $y' = My$ решение $y = Ce^{Mx}$ не может уйти в бесконечность при конечном x.

16. Краевые задачи. До сих пор, для того чтобы выделить частное решение из общего, мы пользовались начальными условиями, согласно которым искомая функция и ее производные задаются при каком-либо одном значении аргумента. Имеются и другие способы выделения частного решения из общего, которые встречаются в практических задачах. Все эти способы объединяет то, что количество дополнительных равенств, накладываемых на искомое решение, должно равняться числу степеней свободы (п. X.2) в общем решении рассматриваемого уравнения, т. е. порядку этого уравнения.

Эти дополнительные равенства в случае уравнения (5) порядка n можно записать в виде

$$G_k[y] = \alpha_k \qquad (k = 1, 2, \ldots, n), \tag{94}$$

где $G_k[y]$ — какая-либо заданная комбинация значений искомой функции $y(x)$ и ее производных при, вообще говоря, различных значениях аргумента (точнее, $G_k[y]$ — это какой-либо заданный *функционал*, см. п. XIV.4), а α_k — заданные числа. Например, в случае начальных условий (11) $G_k[y]$ означает $y^{(k-1)}(x_0)$.

Если известно общее решение (8) заданного уравнения, то для нахождения требуемого частного решения надо выражение для общего решения подставить в условия (94), в результате чего получится система n уравнений с n неизвестными $C_1, C_2, \ldots, C_n$.

Если

$$G_k[C_1y_1 + C_2y_2] \equiv C_1G_k[y_1] + C_2G_k[y_2] \qquad (C_1, C_2 = \text{const}),$$

то условия (94) называются *линейными*; если к тому же все $\alpha_k = 0$, то они называются *линейными однородными*. Если какие-нибудь функции, не обязательно решения дифференциального уравнения, удовлетворяют линейным однородным условиям, то и их любая линейная комбинация тоже удовлетворяет этим условиям. Действительно, если, например, $G_k[y_1] = 0$ и $G_k[y_2] = 0$, то

$$G_k[C_1y_1 + C_2y_2] = C_1G_k[y_1] + C_2G_k[y_2] = C_1 \cdot 0 + C_2 \cdot 0 = 0.$$

Разность двух функций, удовлетворяющих одинаковым неоднородным линейным условиям, удовлетворяет соответствующим однородным условиям (проверьте!).

В дальнейшем мы рассмотрим решение уравнения

$$y'' + p(x)y' + q(x)y = f(x) \qquad (a \leqslant x \leqslant b) \tag{95}$$

при дополнительных условиях

$$y(a) = \alpha_1, \quad y(b) = \alpha_2, \tag{96}$$

хотя *все полученные общие выводы справедливы для линейных дифференциальных уравнений любого порядка n при линейных дополнительных условиях* (94) *любого вида.* Интервал (a, b), а также функции $p(x)$, $q(x)$ и $f(x)$ будем считать конечными, что дает возможность считать любое решение продолженным на весь этот интервал, включая концы (п. 15, свойство 4). Условия вида (96),

наложенные на концах интервала, в котором строится решение, называются *краевыми условиями*, а задача о решении дифференциального уравнения при заданных краевых условиях называется *краевой задачей*.

Для решения краевой задачи мы исходим из вида общего решения уравнения (95)

$$y(x)=Y(x)+C_1z_1(x)+C_2z_2(x) \tag{97}$$

(п. 15, свойство 1), где $Y(x)$ — некоторое частное решение уравнения (95), а z_1 и z_2 — два линейно независимых решения соответствующего однородного уравнения. Подставляя формулу (97) в условия (96), получим два соотношения для нахождения C_1 и C_2:

$$\left.\begin{aligned} C_1z_1(a)+C_2z_2(a)&=\alpha_1-Y(a),\\ C_1z_1(b)+C_2z_2(b)&=\alpha_2-Y(b). \end{aligned}\right\} \tag{98}$$

При решении этой системы двух алгебраических уравнений первой степени с двумя неизвестными могут представиться два случая (п. VI.4 и VI.6).

1. *Основной случай*: определитель системы отличен от нуля В этом случае система (98) имеет вполне определенное решение и потому *уравнение* (95) *при условиях* (96) *имеет одно и только одно решение при любом неоднородном члене* $f(x)$ *и любых числах* α_1, α_2.

2. *Особый случай*: определитель системы равен нулю. В этом случае система (98), как правило, противоречива, но при некоторых правых частях она имеет бесконечное количество решений. Значит, и *уравнение* (95) *при условиях* (96) *при произвольном выборе функции* $f(x)$ *и чисел* α_1, α_2, *как правило, не имеет ни одного решения, однако при некоторых таких выборах задача имеет бесконечное количество решений*. Например, можно проверить, что если $f(x)$ и α_1 уже выбраны, то бесконечное количество решений получится лишь при одном значении α_2, а при остальных значениях задача не будет иметь ни одного решения.

Подчеркнем, что то, какой именно случай имеет место, зависит от вида левых частей уравнения (95) и условий (96).

Согласно п. VI.6, *для того чтобы имел место основной случай, необходимо и достаточно, чтобы соответствующая однородная задача*, в которой положено $f(x)\equiv 0$, $\alpha_1=\alpha_2=0$, *имела только нулевое решение*. В особом случае однородная задача имеет бесконечное количество решений, а если неоднородная задача имеет хотя бы одно решение, то общее решение получится, если к этому частному решению прибавить общее решение соответствующей однородной задачи.

При решении начальной задачи, т. е. задачи Коши, всегда имеет место основной случай, так как такое решение всегда существует и единственно. При решении краевой задачи может представиться и особый случай.

Например, рассмотрим задачу с параметром $\lambda=\text{const}$,

$$y''+\lambda y=f(x)\quad (0\leqslant x\leqslant l),\qquad y(0)=\alpha_1,\qquad y(l)=\alpha_2, \tag{99}$$

причем будем считать сначала, что $\lambda>0$. Тогда линейно независимыми решениями соответствующего однородного дифференциального уравнения служат функции $z_1(x)=\cos\sqrt{\lambda}\,x$, $z_2(x)=\sin\sqrt{\lambda}\,x$ и определитель системы (98) равен

$$\begin{vmatrix} z_1(0) & z_2(0)\\ z_1(l) & z_2(l)\end{vmatrix}=\begin{vmatrix} 1 & 0\\ \cos\sqrt{\lambda}\,l & \sin\sqrt{\lambda}\,l\end{vmatrix}=\sin\sqrt{\lambda}\,l.$$

Приравнивая его нулю, получим значения

$$\lambda=\left(\frac{\pi}{l}\right)^2,\quad \left(\frac{2\pi}{l}\right)^2,\quad \left(\frac{3\pi}{l}\right)^2,\ \ldots, \tag{100}$$

при которых для задачи (99) имеет место особый случай, т. е. нарушается либо существование, либо единственность решения

Набор значений параметра, входящего в формулировку задачи, при которых задача в том или ином смысле вырождается (п. II.8), называется *спектром* этой задачи. Мы предоставляем читателю проверить, что при $\lambda \leqslant 0$ для задачи (99) всегда имеет место основной случай и тем самым набор значений (100) представляет собой ее спектр.

Полученный результат имеет важное приложение к исследованию устойчивости упругого стержня при его сжатии. Пусть однородный (одинаковый по всей длине) упругий невесомый стержень расположен вдоль оси x и сжимается вдоль нее силой P, причем оба конца стержня удерживаются на оси x, но могут свободно вращаться вокруг точек закрепления (рис. 295, *а*). Тогда при достижении силой некоторого *критического значения* $P_{кр}$ стержень выпучивается, принимая положение, изображенное на рис. 295, *б*. Если обозначить через y поперечное отклонение точки стержня от ее исходного положения, то, как доказывается в курсах сопротивления материалов, функция $y(x)$ с достаточной точностью удовлетворяет дифференциальному уравнению и краевым условиям:

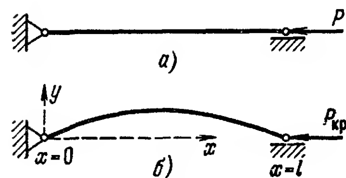


Рис. 295.

$$y'' + \frac{P}{EJ}\,y = 0, \qquad y(0) = y(l) = 0; \tag{101}$$

здесь E и J — так называемые *модуль Юнга* (Т. Юнг, 1773—1829, — английский физик, врач и астроном, один из создателей волновой теории света) и *момент инерции* поперечного сечения стержня. Как вытекает из (100), при

$$\frac{P}{EJ} < \left(\frac{\pi}{l}\right)^2 \tag{102}$$

для задачи (101) имеет место основной случай, т. е. она имеет только нулевое решение; выпучивания не происходит. Как только с увеличением P неравенство (102) переходит в равенство, то наступает особый случай и задача (101) наряду с нулевым решением приобретает решение вида $y = C\sin\frac{\pi}{l}x$, где C — произвольная постоянная. Но тогда стержень ничем не удерживается в прямолинейном состоянии и как угодно малые внешние воздействия могут привести к конечным отклонениям от этого состояния: стержень *теряет устойчивость*. Получающееся выражение для $P_{кр}$

$$P_{кр} = EJ\left(\frac{\pi}{l}\right)^2$$

было найдено Эйлером в 1757 г. Могло бы показаться, что при $P > P_{кр}$ стержень должен опять выпрямиться. Однако это не так. Уравнение (101) описывает отклонение стержня точно лишь в пределе при малых отклонениях, а анализ более точного нелинейного уравнения, справедливого при любых отклонениях, показывает, что при переходе P через $P_{кр}$ наряду с неустойчивой прямолинейной возникает искривленная форма равновесия, которая и является устойчивой. С ростом P кривизна этой формы быстро возрастает и стержень разрушается.

К решению неоднородного уравнения при однородных краевых условиях

$$y'' + p(x)\,y' + q(x)\,y = f(x) \quad (a \leqslant x \leqslant b), \qquad y(a) = 0, \qquad y(b) = 0, \tag{103}$$

в основном (неособом) случае можно применить функцию влияния (п. XIV.26). Действительно, функцию $f(x)$ можно истолковать как «внешнее воздействие»,

а $y(x)$ — как его результат, т. е. $y(x)=\tilde{f}(x)$. При этом имеет место принцип суперпозиции (почему?).

В соответствии с п. XIV.26, если через $G(x;\ \xi)$ обозначить решение задачи (103), в которой вместо $f(x)$ взята дельта-функция $\delta(x-\xi)$, то при произвольной функции $f(x)$ решение задачи (103) получится по формуле

$$y(x)=\int_a^b f(\xi)\,G(x;\ \xi)\,d\xi. \qquad (104)$$

Приведем простой пример. Пусть рассматривается задача

$$y''=f(x)\quad (0\leqslant x\leqslant l),\qquad y(0)=y(l)=0. \qquad (105)$$

Если взамен $f(x)$ поставить $\delta(x-\xi)$, то при $0\leqslant x<\xi$ и при $\xi<x\leqslant l$ получаем просто $y''=0$, т. е. решение

$$y=ax+b\quad (0\leqslant x<\xi),\qquad y=cx+d\quad (\xi<x\leqslant l),$$

где a, b, c, d — какие-то постоянные. Применение краевых условий показывает, что $b=0$ и $cl+d=0$, т. е.

$$y=ax\quad (0\leqslant x<\xi),\qquad y=c(x-l)\quad (\xi<x\leqslant l). \qquad (106)$$

Если равенство $y''=\delta(x-\xi)$ проинтегрировать от $x=\xi-0$ до $x=\xi+0$, то получится, что $y'(\xi+0)-y'(\xi-0)=1$; кстати, для левой части уравнения (103) получился бы такой же результат, так как интегрирование конечной функции по отрезку нулевой длины дает нуль. При вторичном интегрировании дельта-функции получается уже непрерывная функция (п XIV.25), так что $y(\xi-0)=y(\xi+0)$, и из (106) получаем $c-a=1$, $a\xi=c(\xi-l)$, откуда

$$a=-\frac{l-\xi}{l},\qquad c=\frac{\xi}{l}.$$

Подставляя в (106), находим функцию влияния для задачи (105):

$$G(x,\ \xi)=\begin{cases} -\dfrac{(l-\xi)\,x}{l} & (0\leqslant x<\xi);\\ -\dfrac{\xi\,(l-x)}{l} & (\xi<x\leqslant l).\end{cases}$$

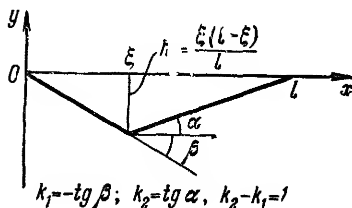


Рис. 296.

Эта функция изображена на рис. 296. В силу формулы (104) получаем решение задачи (105) при любой функции $f(x)$:

$$y=\int_0^l G(x;\ \xi)\,f(\xi)\,d\xi=\int_0^x G(x;\ \xi)\,f(\xi)\,d\xi+\int_x^l G(x;\ \xi)\,f(\xi)\,d\xi=$$

$$=-\frac{(l-x)}{l}\int_0^x \xi f(\xi)\,d\xi-\frac{x}{l}\int_x^l (l-\xi)\,f(\xi)\,d\xi.$$

§ 5. Линейные уравнения с постоянными коэффициентами

Линейные уравнения с постоянными коэффициентами составляют важнейший класс уравнений, интегрирование которых сравнительно легко доводится до конца.

17. Однородные уравнения. Рассмотрим для определенности уравнение третьего порядка

$$z''' + a_1 z'' + a_2 z' + a_3 z = 0 \qquad (a_1,\ a_2,\ a_3 = \text{const}). \tag{107}$$

Эйлер предложил искать частные решения этого уравнения в форме

$$z = e^{px}, \tag{108}$$

где p — постоянная, которую нужно подобрать.

Подставляя (108) в (107), получим, что

$$e^{px}(p^3 + a_1 p^2 + a_2 p + a_3) = 0,$$

и так как первый множитель отличен от нуля, то

$$p^3 + a_1 p^2 + a_2 p + a_3 = 0. \tag{109}$$

Итак, *для того чтобы функция* (108) *удовлетворяла уравнению* (107), *необходимо и достаточно, чтобы* p *удовлетворяло уравнению* (109). *Алгебраическое* относительно p *уравнение* (109) называется *характеристическим уравнением* для уравнения (107), а левая часть уравнения (109) называется *характеристическим многочленом* для уравнения (107). Степень характеристического уравнения равна порядку соответствующего дифференциального уравнения.

Уравнение (109) имеет три корня (п. VIII.8): p_1, p_2, p_3. При этом могут быть различные случаи.

1. *Пусть все корни вещественные и простые*, т. е. различные. Тогда в силу формулы (108) мы имеем три частных решения уравнения (107)

$$z_1 = e^{p_1 x}, \quad z_2 = e^{p_2 x}, \quad z_3 = e^{p_3 x}.$$

Так как они являются независимыми, т. е. ни одно из них не равно линейной комбинации остальных, то на основе п. 14 общее решение уравнения (107) имеет вид

$$z = C_1 e^{p_1 x} + C_2 e^{p_2 x} + C_3 e^{p_3 x}. \tag{110}$$

2. *Пусть все корни простые, но среди них имеются мнимые.* Тогда в правой части формулы (110) оказывается комплексная функция от вещественного аргумента (п. VIII.6). Но вся теория линейных уравнений (§ 4) автоматически распространяется на случай, когда все коэффициенты и решения являются такими функциями. Поэтому *и при указанных корнях уравнения* (109) *можно пользоваться формулой* (110); конечно, тогда произвольные постоянные будут, вообще говоря, комплексными.

Однако если все рассмотрения производятся над вещественными функциями, то часто предпочитают и ответ получить в вещественной форме. Для этого можно воспользоваться следующим замечанием: *если линейное однородное уравнение с вещественными коэффициентами имеет комплексное частное решение, то его вещественная и мнимая*

части также являются решениями того же уравнения. Действительно, если $\mathrm{L}[y_1+iy_2]=0$ (см. обозначение в п. 14), то $\mathrm{L}[y_1]+i\mathrm{L}[y_2]=0$, откуда $\mathrm{L}[y_1]=0$ и $\mathrm{L}[y_2]=0$ (почему?).

Значит, если коэффициенты уравнения (107) вещественные и оно имеет частное решение

$$e^{(r+is)x}=e^{rx}\cos sx+ie^{rx}\sin sx$$

(см. формулу (VIII.12)), то функции

$$e^{rx}\cos sx, \quad e^{rx}\sin sx \tag{111}$$

также служат решениями уравнения (107). Если вспомнить (п. VIII.8), что у алгебраического уравнения с вещественными коэффициентами сопряженные корни присутствуют парами, то получаем, что в рассматриваемом случае 2 корни уравнения (109) имеют вид

$$p_1=r+is, \quad p_2=r-is \qquad (p_3 \text{ — вещественное}),$$

и потому решение можно вместо (110) записать в вещественной форме:

$$z=C_1e^{rx}\cos sx+C_2e^{rx}\sin sx+C_3e^{p_3x}. \tag{112}$$

Например, для уравнения свободных колебаний (88) получаем характеристическое уравнение $p^2+\omega_0^2=0$ с корнями $p_{1,2}=\pm i\omega_0=$ $=0\pm i\omega_0$ и аналогично формуле (112) пишем общее решение

$$z=C_1e^{0t}\cos\omega_0t+C_2e^{0t}\sin\omega_0t,$$

т. е. как раз решение (89).

Формулу (112) иногда записывают в ином виде, преобразовав

$$C_1\cos sx+C_2\sin sx=M\sin(sx+\alpha),$$

для чего надо положить

$$C_1=M\sin\alpha, \quad C_2=M\cos\alpha, \quad M=\sqrt{C_1^2+C_2^2}, \quad \operatorname{tg}\alpha=\frac{C_1}{C_2}$$

(ср. п. I.29). Тогда взамен (112) получим

$$z=Me^{rx}\sin(sx+\alpha)+C_3e^{p_3x}, \tag{113}$$

где произвольными постоянными являются уже M, α и C_3.

3. *Пусть среди корней характеристического уравнения* (109) *имеются кратные, например,* $p_2=p_1$, $p_3\neq p_1$. Тогда формула (110), конечно, не даст общего решения и в виде (108) мы получим лишь два решения, e^{p_1x} и e^{p_3x}.

Чтобы найти третье решение, рассмотрим сначала случай, когда $p_2=p_1+\Delta p$, причем $|\Delta p|$ мало́. Тогда уравнение (107) наряду с решением e^{p_1x} имеет решение

$$e^{p_2x}=e^{p_1x}e^{\Delta p\cdot x}=e^{p_1x}\left(1+\Delta p\cdot x+\frac{(\Delta p)^2x^2}{2!}+\ldots\right)$$

(см. разложение (IV.55)), а потому служат решениями и их линейные комбинации

$$e^{p_2 x}-e^{p_1 x}=e^{p_1 x}\left(\Delta p\cdot x+\frac{(\Delta p)^2x^2}{2!}+\ldots\right),$$

$$\frac{e^{p_2 x}-e^{p_1 x}}{\Delta p}=e^{p_1 x}\left(x+\frac{\Delta p\cdot x^2}{2!}+\ldots\right). \quad (114)$$

Это деление на Δp дает возможность перейти к пределу при $\Delta p\to 0$. Тогда в правой части все члены, содержащие Δp, отпадут, и потому при $\Delta p=0$, т. е. $p_2=p_1$, решением служит функция $xe^{p_1 x}$.

Значит в рассматриваемом случае решением уравнения (107) будет

$$z=C_1e^{p_1 x}+C_2xe^{p_1 x}+C_3e^{p_3 x}.$$

Подобным образом в случае $p_1=p_2=p_3$ частными решениями уравнения (107) наряду с $e^{p_1 x}$ служат функции $xe^{p_1 x}$ и $x^2e^{p_1 x}$; при доказательстве этого надо взамен (114) рассмотреть вторую разделенную разность (см. п. V.7). Поэтому в данном случае общее решение имеет вид

$$z=C_1e^{p_1 x}+C_2xe^{p_1 x}+C_3x^2e^{p_1 x}.$$

Рассмотрение уравнений любого порядка проходит аналогично. Если какой-либо корень p характеристического уравнения имеет кратность k, то функции

$$e^{px},\quad xe^{px},\quad \ldots,\quad x^{k-1}e^{px}$$

будут частными решениями рассматриваемого дифференциального уравнения. Если какая-либо пара корней $r\pm is$ имеет кратность k, то частными решениями будут функции

$$e^{rx}\cos sx,\quad e^{rx}\sin sx,\quad xe^{rx}\cos sx,\quad xe^{rx}\sin sx,\ \ldots$$
$$\ldots,\ x^{k-1}e^{rx}\cos sx,\ x^{k-1}e^{rx}\sin sx.$$

Итак, практическая трудность при решении линейного однородного уравнения с постоянными коэффициентами состоит единственно в решении соответствующего характеристического уравнения, что делается по методам п. V.2—5 и VIII.9.

В качестве примера рассмотрим свободные колебания материальной точки при линейном законе упругости и при дополнительном *вязком трении*, пропорциональном первой степени скорости. В этом случае в правой части уравнения (3) надо добавить слагаемое $-f\frac{dy}{dt}$, где f — *коэффициент трения*. После переноса всех членов налево и деления на M получим аналогично (88) уравнение

$$z''+2hz'+\omega_0^2z=0,\quad \text{где}\quad 2h=\frac{f}{M},\quad \omega_0^2=\frac{k}{M}. \quad (115)$$

При решении характеристического уравнения

$$p^2+2hp+\omega_0^2=0 \tag{116}$$

возникают два основных случая. Если $h<\omega_0$, т. е. если трение сравнительно мало, уравнение (116) имеет решение

$$p_{1,2}=-h\pm\sqrt{h^2-\omega_0^2}=-h\pm i\sqrt{\omega_0^2-h^2},$$

а потому общее решение уравнения (115) имеет вид, подобный (113),

$$z(t)=Me^{-ht}\sin(\omega t+\alpha),\quad \text{где}\quad \omega=\sqrt{\omega_0^2-h^2}.$$

Мы видим, что наличие небольшого трения делает колебания затухающими по экспоненциальному закону (множитель e^{-ht}) и уменьшает частоту (так как $\omega<\omega_0$). График решения показан на рис. 297. Нули решения определяются множителем $\sin(\omega t+\alpha)$ и потому находятся на равном расстоянии друг от друга. Через каждый промежуток времени $T=\frac{2\pi}{\omega}$, когда синус повторяет свои значения, из e^{-ht} выделяется множитель $e^{-h\frac{2\pi}{\omega}}$, из-за чего и происходит затухание. Значение T часто называется «периодом» колебания, хотя $z(t)$ здесь непериодическая функция, так как на каждом следующем «периоде» колебание, не меняясь по форме, уменьшается по размаху в одно и то же число раз.

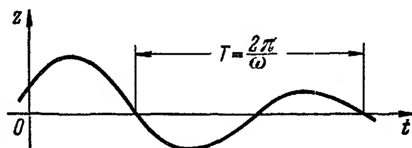


Рис 297.

Если $h>\omega_0$, т. е. если трение сравнительно велико, то уравнение (116) имеет вещественные корни, а уравнение (115) имеет общее решение

$$z(t)=C_1e^{p_1t}+C_2e^{p_2t}=C_1e^{-\left(h-\sqrt{h^2-\omega_0^2}\right)t}+C_2e^{-\left(h+\sqrt{h^2-\omega_0^2}\right)t}.$$

При больших t здесь существенно только первое слагаемое (почему?), т е. мы получаем затухание по экспоненциальному закону без колебаний; это — так называемое *апериодическое затухание.*

Теоретически возможен «пограничный случай» $h=\omega_0$. Тогда уравнение (116) имеет двойной корень. Предоставим читателю убедиться в том, что и здесь получится апериодическое затухание.

18. Неоднородные уравнения с правыми частями специального вида. Рассмотрим теперь линейное неоднородное уравнение с постоянными коэффициентами, например, третьего порядка:

$$y'''+a_1y''+a_2y'+a_3y=f(x)\qquad (a_1,\ a_2,\ a_3=\text{const}). \tag{117}$$

Поскольку соответствующее однородное уравнение всегда решается (п. 17), то в силу п. 15 нам остается лишь найти какое-либо частное решение уравнения (117). Для правой части общего

вида это делается по методу вариации произвольных постоянных (п. 15). Но для важного довольно широкого класса правых частей специального вида частное решение можно найти значительно быстрее по методу неопределенных коэффициентов.

Рассмотрим сначала уравнение

$$y''' + a_1 y'' + a_2 y' + a_3 y = Ke^{\lambda x} \qquad (K,\ \lambda = \text{const}). \tag{118}$$

Естественно искать частное решение этого уравнения в форме

$$y = Ae^{\lambda x}, \tag{119}$$

где постоянная A пока неизвестна. Подстановка в (118) даст

$$A\lambda^3 e^{\lambda x} + Aa_1\lambda^2 e^{\lambda x} + Aa_2\lambda e^{\lambda x} + Aa_3 e^{\lambda x} = Ke^{\lambda x},$$

или, после сокращения,

$$A = \frac{K}{\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3}, \qquad y = \frac{K}{P(\lambda)} e^{\lambda x}, \tag{120}$$

где через $P(\lambda)$ обозначен характеристический многочлен.

Полученный результат годится, если $P(\lambda) \neq 0$, т. е. если λ не является корнем характеристического уравнения. Если $P(\lambda) = 0$, то функция (119) удовлетворяет однородному уравнению (107), т. е. уравнению (118) удовлетворить в такой форме невозможно.

Пусть $P(\lambda) = 0$, $P'(\lambda) \neq 0$, т. е. λ является простым (п. VIII.8) корнем характеристического уравнения. Тогда рассуждаем подобно тому, как мы в п. 17 рассматривали случай кратных корней. Заменим в правой части (118) λ на $\lambda_1 = \lambda + \alpha$, где $|\alpha| \neq 0$, но мало́. Тогда в силу формулы (120), так как λ_1 уже не будет корнем характеристического уравнения, уравнение (118) будет обладать частным решением

$$\frac{K}{P(\lambda_1)} e^{\lambda_1 x} = \frac{K}{P(\lambda+\alpha)} e^{(\lambda+\alpha)x} = \frac{K}{P(\lambda+\alpha)} e^{\lambda x}\left(1 + \alpha x + \frac{\alpha^2 x^2}{2!} + \ldots\right) =$$
$$= \frac{K}{P(\lambda+\alpha)} e^{\lambda x} + K\frac{\alpha}{P(\lambda+\alpha)} e^{\lambda x}\left(x + \frac{\alpha x^2}{2!} + \ldots\right).$$

Однако первое из полученных слагаемых удовлетворяет соответствующему однородному уравнению; значит, второе слагаемое также является частным решением уравнения (118), в котором λ пока еще заменено на λ_1 (почему?). Если в этом втором слагаемом перейти к пределу при $\alpha \to 0$, вычислив $\lim\limits_{\alpha \to 0} \frac{\alpha}{P(\lambda+\alpha)}$ по правилу Лопиталя, то мы в пределе получим частное решение уравнения (118) при исходном значении λ (проверьте!):

$$y = \frac{K}{P'(\lambda)} xe^{\lambda x}. \tag{121}$$

Подобным образом, если λ является двойным корнем характеристического уравнения, то частное решение уравнения (118) имеет вид

$$y = \frac{K}{P''(\lambda)} x^2 e^{\lambda x}$$

и т. д.

С помощью аналогичного, но более громоздкого рассуждения можно доказать, что уравнение

$$y''' + a_1 y'' + a_2 y' + a_3 y = Q_m(x) e^{\lambda x}, \tag{122}$$

где $Q_m(x)$ — заданный многочлен степени m, если λ не является корнем характеристического уравнения, обладает частным решением вида

$$y = R_m(x) e^{\lambda x}, \tag{123}$$

где $R_m(x)$ — некоторый другой многочлен степени m. Его можно найти по методу неопределенных коэффициентов, т. е. написать его сначала с буквенными коэффициентами, подставить в (122) и найти эти коэффициенты из условия тождественного равенства левой и правой частей. Если же λ является корнем характеристического уравнения кратности k, то имеется частное решение вида

$$y = x^k R_m(x) e^{\lambda x}.$$

При этом не исключен случай $\lambda = 0$, когда в правой части уравнения (122) стоит «чистый» многочлен.

Можно также рассмотреть уравнение

$$y''' + a_1 y'' + a_2 y' + a_3 y = Q_m(x) e^{\mu x} \cos \nu x.$$

Так как правую часть можно переписать в виде (п. VIII.4)

$$Q_m(x) e^{\mu x} \frac{e^{i\nu x} + e^{-i\nu x}}{2} = \frac{Q_m(x)}{2} e^{(\mu + i\nu) x} + \frac{Q_m(x)}{2} e^{(\mu - i\nu) x},$$

то в силу (123), если $\lambda = \mu \pm i\nu$ не является корнем характеристического уравнения, частное решение можно искать в виде

$$\begin{aligned} y = R_m(x) e^{(\mu+i\nu) x} + \tilde{R}_m(x) e^{(\mu-i\nu) x} = \\ = R_m(x) e^{\mu x} (\cos \nu x + i \sin \nu x) + \tilde{R}_m(x) e^{\mu x} (\cos \nu x - i \sin \nu x) = \\ = [R_m(x) + \tilde{R}_m(x)] e^{\mu x} \cos \nu x + [iR_m(x) - i\tilde{R}_m(x)] e^{\mu x} \sin \nu x. \end{aligned}$$

Вводя новые обозначения, получим частное решение вида

$$y = T_m(x) e^{\mu x} \cos \nu x + S_m(x) e^{\mu x} \sin \nu x, \tag{124}$$

где $T_m(x)$ и $S_m(x)$ — многочлены степени m, которые можно найти по методу неопределенных коэффициентов. В виде (124) ищутся

частные решения уравнений и с правыми частями

$$Q_m(x)\, e^{\mu x} \sin \nu x, \quad Q_m(x)\, e^{\mu x} \cos(\nu x + \alpha),$$
$$Q_m(x)\, e^{\mu x} \sin(\nu x + \alpha) \qquad (\alpha = \text{const}).$$

Если $\lambda = \mu \pm i\nu$ является корнем характеристического уравнения кратности k, то правую часть формулы (124) надо умножить еще на x^k.

Рассмотрим, например, уравнение вынужденных колебаний (87) при синусоидальном внешнем воздействии с частотой ω:

$$y'' + \omega_0^2 y = K \sin \omega t. \tag{125}$$

Согласно формуле (124), если $\lambda = \pm i\omega$ не является корнем характеристического уравнения, т. е. если $\omega \neq \omega_0$, то частное решение надо искать в виде

$$y = A \cos \omega t + B \sin \omega t.$$

Подстановка в уравнение (125) дает

$$-A\omega^2 \cos \omega t - B\omega^2 \sin \omega t + A\omega_0^2 \cos \omega t + B\omega_0^2 \sin \omega t = K \sin \omega t.$$

Так как это равенство должно быть тождеством, то

$$-A\omega^2 + A\omega_0^2 = 0, \quad -B\omega^2 + B\omega_0^2 = K,$$

откуда

$$A = 0, \qquad B = \frac{K}{\omega_0^2 - \omega^2}, \quad y = \frac{K}{\omega_0^2 - \omega^2} \sin \omega t. \tag{126}$$

Общее решение уравнения (125) получится, если добавить общее решение (89) соответствующего однородного уравнения. Итак, если частота внешнего воздействия не равна собственной частоте колебаний, то получается наложение двух гармонических колебаний. Одно, обычно называемое *вынужденным*, происходит с частотой внешнего воздействия и имеет вполне определенные амплитуду и начальную фазу; другое происходит с собственной частотой, и его амплитуда и начальная фаза зависят от начальных данных.

Из формулы (126) видно, что если ω близко к ω_0, то амплитуда вынужденного колебания становится очень большой. Если же $\omega = \omega_0$, то согласно общей теории частное решение уравнения (125) надо искать в форме $y = At \cos \omega_0 t + B t \sin \omega_0 t$.

Подсчет дает (проверьте!)

$$y = -\frac{K}{2\omega_0} t \cos \omega_0 t;$$

это можно вывести также из (126) аналогично формуле (121).

Мы видим, что если частота внешнего воздействия равна собственной частоте колебаний, то амплитуда вынужденного колебания возрастает по линейному закону. Это важное явление хорошо известно в физике и технике и называется *резонансом*.

19. Уравнение Эйлера. Так называется линейное уравнение вида

$$(ax+b)^n y^{(n)}+a_1(ax+b)^{n-1}y^{(n-1)}+\ldots+a_{n-1}(ax+b)y'+a_ny=f(x),$$

в котором все коэффициенты $a_1, a_2, \ldots, a_n$ постоянны

Уравнение Эйлера легко приводится к линейному уравнению с постоянными коэффициентами при помощи замены независимой переменной

$$|ax+b|=e^t, \qquad t=\ln|ax+b|.$$

Будем для простоты считать, что $ax+b>0$, а уравнение однородно и имеет второй порядок:

$$(ax+b)^2 y''+a_1(ax+b)y'+a_2y=0. \tag{127}$$

После замены независимой переменной получим

$$ax+b=e^t, \qquad t=\ln(ax+b), \tag{128}$$

$$y'=\frac{dy}{dx}=\frac{dy}{dt}\cdot\frac{dt}{dx}=\frac{dy}{dt}\cdot ae^{-t};$$

$$y''=\frac{dy'}{dx}=\frac{dy'}{dt}\cdot\frac{dt}{dx}=\left(\frac{d^2y}{dt^2}ae^{-t}-\frac{dy}{dt}ae^{-t}\right)ae^{-t}.$$

Подставим эти результаты в уравнение (127):

$$a^2\left(\frac{d^2y}{dt^2}-\frac{dy}{dt}\right)+a_1a\frac{dy}{dt}+a_2y=0.$$

Это — уравнение с постоянными коэффициентами, которое надо решать по методам п. 17, т. е. положить

$$y=e^{pt}, \tag{129}$$

решить характеристическое уравнение и т. д., после чего вернуться от t к x.

Можно не делать замены (128), заметив, что из (128) и (129) следует

$$y=(ax+b)^p. \tag{130}$$

Поэтому можно путем непосредственной подстановки в (127) искать частные решения вида (130), причем для нахождения p получится характеристическое уравнение, степень которого равна порядку уравнения (127). Надо только иметь в виду, что при наличии кратных корней характеристического уравнения, помимо решений вида (130), уравнение (127) будет обладать решением вида

$$y=te^{pt}=(ax+b)^p\ln(ax+b)$$

и т. д., в зависимости от кратности корня (п. 17, случай 3).

20. Операторы и операторное решение уравнений. В п. XIV.26 было определено понятие оператора и были приведены примеры операторов, в том числе оператор дифференцирования D; в п 14 был введен *дифференциальный оператор* L. Другими распространенными операторами являются: *оператор сдвига* T и *оператор образования разности* Δ, действующие по формулам

$$\mathrm{T}f(x)=f(x+h); \qquad \Delta f(x)=f(x+h)-f(x) \tag{131}$$

при заданном шаге h; оператор C умножения на заданное число C, в том числе *единичный оператор* 1, оставляющий функции без изменения, и *нулевой оператор* 0, переводящий все функции в тождественно нулевую функцию; оператор умножения на какую-либо заданную функцию и т. п.

Операторы можно складывать друг с другом и умножать на числа по естественному правилу: если A и B — операторы, а α — число, то по определению

$$(\mathrm{A}+\mathrm{B})f=\mathrm{A}f+\mathrm{B}f, \qquad (\alpha\mathrm{A})f=\alpha(\mathrm{A}f).$$

Например, из равенств (131) видно, что

$$\Delta = \mathrm{T} - 1, \qquad \mathrm{T} = 1 + \Delta.$$

При этом выполняются все аксиомы линейных действий (п. VII.17).

В соответствии с п. XI.6 операторы можно умножать друг на друга, что дает новый оператор, действующий по следующему правилу:

$$(\mathrm{AB})\, f = \mathrm{A}\,(\mathrm{B}f),$$

т. е. на функцию f действует сначала оператор B, а затем на результат — оператор A. Нетрудно проверить правила

$$\mathrm{A}\,(\mathrm{BC}) = (\mathrm{AB})\,\mathrm{C}, \quad (\alpha\mathrm{A} + \beta\mathrm{B})\,\mathrm{C} = \alpha\mathrm{AC} + \beta\mathrm{BC} \qquad (\alpha,\ \beta = \mathrm{const}). \tag{132}$$

Однако далеко не всегда AB = BA, т. е. результат выполнения двух операций может зависеть от порядка действий. Если все же AB = BA, то операторы A и B называются *перестановочными* (*коммутирующими*) друг с другом. Например, все приведенные выше операторы D, T, Δ, C перестановочны друг с другом, так как

$$\mathrm{DT}f(x) = \mathrm{D}\,(\mathrm{T}f(x)) = \\ = \mathrm{D}\,(f(x+h)) = f'(x+h), \ \ \mathrm{TD}f(x) = \mathrm{T}\,(f'(x)) = f'(x+h) \text{ и т. п.}$$

С другой стороны, операторы дифференцирования и умножения на *функцию* неперестановочны (продумайте это!).

Первое свойство (132) дает возможность взамен написанных там выражений писать просто ABC; получается оператор, состоящий в последовательном применении действий C, B и A. Аналогично определяется оператор ABCD и т. д. Если взять множители равными, то получатся *степени оператора*: A^2, A^3, A^4 и т. д., которые означают результат повторения оператора A. Например, $\mathrm{D}^2 f = f''$; $\Delta^2 f$ — это вторая разность (п. V.7) и т. п.

Оператор A называется *линейным* (п. XI.6), если

$$\mathrm{A}\,(f_1 + f_2) = \mathrm{A}\,(f_1) + \mathrm{A}\,(f_2) \quad \text{и} \quad \mathrm{A}\,(\alpha f) = \alpha\mathrm{A}\,(f) \qquad (\alpha = \mathrm{const}). \tag{133}$$

Первое свойство можно истолковать как принцип суперпозиции, а второе возможно вывести из первого (п. XIV 26). Поэтому даже если явный вид оператора неизвестен, то при выполнении принципа суперпозиции можно заключить о линейности этого оператора, что дает возможность сделать некоторые полезные выводы, например построить функцию влияния (см. п. XIV.26).

Оба свойства (133) можно написать вместе так:

$$\mathrm{A}\,(\alpha f_1 + \beta f_2) = \alpha\mathrm{A}f_1 + \beta\mathrm{A}f_2 \qquad (\alpha,\ \beta = \mathrm{const}).$$

Для линейного оператора A нетрудно проверить свойство

$$\mathrm{A}\,(\alpha\mathrm{B} + \beta\mathrm{C}) = \alpha\mathrm{AB} + \beta\mathrm{AC}; \tag{134}$$

для этого надо обе части применить к любой функции f и показать, что получится одно и то же, именно, $\alpha\mathrm{A}\,(\mathrm{B}f) + \beta\mathrm{A}\,(\mathrm{C}f)$.

Все операторы, приведенные выше в качестве примеров, линейные, так как производная суммы равна сумме производных и т. п. Нелинейными являются, например, операторы возведения функции в квадрат или образования абсолютной величины (почему?) и т. п. Для линейных операторов в силу (132) и (134) при линейных действиях и умножении можно пользоваться обычными правилами школьной алгебры, следя за порядком множителей; например, $(\mathrm{A} + \mathrm{B})^2 = (\mathrm{A} + \mathrm{B})(\mathrm{A} + \mathrm{B}) = \mathrm{A}^2 + \mathrm{AB} + \mathrm{BA} + \mathrm{B}^2$ и т. п.

Если к тому же операторы перестановочны, то и порядок множителей несуществен, т. е. $(\mathrm{A} + \mathrm{B})^2 = \mathrm{A}^2 + 2\mathrm{AB} + \mathrm{B}^2$ и т. п.

Можно рассматривать и степенные ряды (п. IV.16) от операторов; например,

$$e^{A}=1+A+\frac{A^2}{2!}+\frac{A^3}{3!}+\ldots \qquad (135)$$

и т. п. Применять такой ряд можно, как и любой оператор, вообще говоря, не ко всем функциям, а только к тем, для которых он имеет смысл. В примере (135) это значит, что чем больше взять членов, тем точнее должен получиться результат; совершенно точный результат теоретически получается лишь в пределе, а практически — при достаточно большом числе членов ряда.

Ряд Тейлора

$$f(x+h)=f(x)+\frac{f'(x)}{1!}h+\frac{f''(x)}{2!}h^2+\ldots$$

(п. IV.16) можно записать в виде

$$Tf=\left(1+\frac{D}{1!}h+\frac{D^2}{2!}h^2+\ldots\right)f=e^{hD}f,$$

откуда мы видим связь между операторами T, Δ и D:

$$T=e^{hD}, \qquad \Delta=e^{hD}-1.$$

Обратная формула

$$D=\frac{1}{h}\ln T=\frac{1}{h}\ln(1+\Delta)=\frac{1}{h}\left(\Delta-\frac{\Delta^2}{2}+\frac{\Delta^3}{3}-\ldots\right)$$

(разложение логарифма по формуле (IV.61)) — это не что иное, как формула (V.32) для численного дифференцирования.

Можно рассматривать *операторное уравнение*

$$Ay=f, \qquad (136)$$

где функция f задана, а функция y ищется. Если решение имеется, то его естественно обозначить $y=A^{-1}f$. Если оператор A линейный, то и уравнение (136) называется линейным. На линейные уравнения немедленно распространяются свойства 1—3 п. 14 и свойство 1 п. 15; однако надо иметь в виду, что бывают случаи, когда однородное уравнение имеет бесконечное количество линейно независимых решений, а также случаи, когда неоднородное уравнение не имеет ни одного решения.

Покажем применение оператора дифференцирования к решению (*операторный метод решения*) линейного дифференциального уравнения с постоянными коэффициентами (117). Уравнение перепишем в виде

$$(D^3+a_1D^2+a_2D+a_3)\,y=f(x).$$

В левой части в скобках стоит линейный дифференциальный оператор третьего порядка с постоянными коэффициентами. По правилам алгебры (п. VIII.8) разлагаем его на множители:

$$(D-p_1)(D-p_2)(D-p_3)\,y=f(x), \qquad (137)$$

где p_1, p_2, p_3 — корни характеристического уравнения (п. 17). Заметив, что

$$D(e^{-px}y)=(e^{-px}y)'=-pe^{-px}y+e^{-px}y'=e^{-px}(y'-py)=e^{-px}(D-p)\,y$$

и потому

$$(D-p)\,y=e^{px}D(e^{-px}y),$$

перепишем уравнение (137) в виде

$$e^{p_1x}D(e^{-p_1x}\,e^{p_2x}D(e^{-p_2x}\,e^{p_3x}D(e^{-p_3x}y)))=f(x).$$

Перенося множители из левой части в правую и пользуясь тем, что равенство $\mathrm{D}y = z$ равносильно $y = \int z\,dx$, получим общее решение исходного уравнения

$$y = e^{p_3 x}\int e^{(p_2 - p_3)\,x}\left(\int e^{(p_1 - p_2)\,x}\left(\int e^{-p_1 x} f(x)\,dx\right)dx\right)dx.$$

Конечно, результат получится тот же, что в пп. 15 и 18, хотя подход здесь несколько иной; в более сложных задачах применение операторов может принести существенную пользу. Отметим, что для линейных дифференциальных операторов с переменными коэффициентами и тем более для нелинейных операторов такое простое разложение (*факторизацию*) оператора на произведение нескольких операторов более низкого порядка эффективно удается осуществить лишь в очень редких случаях.

Приведем еще один простой пример. Пусть надо найти решение уравнения

$$y'' + \omega^2 y = f(x),$$

причем все величины считаются вещественными. Пишем

$$(\mathrm{D}^2 + \omega^2)\,y = f(x), \quad (\mathrm{D} - i\omega)(\mathrm{D} + i\omega)\,y = f(x),$$

$$e^{i\omega x}\mathrm{D}\left(e^{-i\omega x}(\mathrm{D} + i\omega)\,y\right) = f(x);$$

$$(\mathrm{D} + i\omega)\,y = e^{i\omega x}\int e^{-\,\omega x} f(x)\,dx,$$

$$\omega y = \mathrm{Im}\left(e^{i\omega x}\int e^{-i\omega x} f(x)\,dx\right)$$

(Im означает мнимую часть, см. пп. VIII.1 и VIII.6). Окончательно,

$$y = \frac{1}{\omega}\,\mathrm{Im}\left(e^{i\omega x}\int e^{-i\omega x} f(x)\,dx\right).$$

§ 6. Системы линейных уравнений

21. Системы линейных уравнений. Рассмотрим для определенности *линейную однородную систему* трех уравнений первого порядка с тремя искомыми функциями $y(x)$, $z(x)$ и $u(x)$, разрешенных относительно производных от этих функций:

$$\left.\begin{aligned} y' &= a_1(x)\,y + b_1(x)\,z + c_1(x)\,u, \\ z' &= a_2(x)\,y + b_2(x)\,z + c_2(x)\,u, \\ u' &= a_3(x)\,y + b_3(x)\,z + c_3(x)\,u. \end{aligned}\right\} \tag{138}$$

Напомним, что систему уравнений любого порядка легко преобразовать в систему первого порядка (см. п. 11), а разрешение системы относительно производных осуществляется алгебраически, без решения самих дифференциальных уравнений

Так как от системы (138) легко перейти к равносильному уравнению третьего порядка (п. 11), которое также получается линейным и однородным, то все свойства линейных однородных уравнений (п. 14) распространяются на систему (138). При этом суммой двух решений $y = y_1$, $z = z_1$, $u = u_1$ и $y = y_2$, $z = z_2$, $u = u_2$ считается решение $y = y_1 + y_2$, $z = z_1 + z_2$, $u = u_1 + u_2$, а произведением решения $y = y_1$, $z = z_1$, $u = u_1$ на число C считается решение $y = Cy_1$, $z = Cz_1$, $u = Cu_1$, т. е. линейные действия над решениями осуществляются так же, как над векторами (п. VII.10).

В частности, *общее решение системы* (138) *имеет вид*

$$\left.\begin{aligned} y&=C_1y_1+C_2y_2+C_3y_3,\\ z&=C_1z_1+C_2z_2+C_3z_3,\\ u&=C_1u_1+C_2u_2+C_3u_3,\end{aligned}\right\}\tag{139}$$

где C_1, C_2, C_3 — произвольные постоянные, а (y_1, z_1, u_1), (y_2, z_2, u_2) и (y_3, z_3, u_3) — три *линейно независимых решения* системы (138), т е таких, что ни одно из них не является линейной комбинацией остальных

Остановимся на свойстве 4 п. 14 Если известно ненулевое решение (y_1, z_1, u_1) системы (138), то, сделав подстановку $y=y_1\bar{y}$, $z=z_1\bar{z}$, $u=u_1\bar{u}$, а затем $\bar{y}=\bar{u}+v$, $\bar{z}=\bar{u}+w$, нетрудно получить систему *двух* уравнеиий первого порядка с *двумя* неизвестиыми функциями, $v(x)$ и $w(x)$, из которой $\bar{u}$ находится с помощью одного интегрирования

Все это мы предоставим проверить желающим

На *линейные неоднородные* системы вида

$$\left.\begin{aligned} y'&=a_1(x)y+b_1(x)z+c_1(x)u+f_1(x),\\ z'&=a_2(x)y+b_2(x)z+c_2(x)u+f_2(x),\\ u'&=a_3(x)y+b_3(x)z+c_3(x)u+f_3(x)\end{aligned}\right\}\tag{140}$$

также распространяются все свойства, указанные в п. 15 В частности, метод вариации произвольных постоянных (свойство 3) выглядит так. Пусть известно общее решение (139) соответствующей однородной системы (138). Тогда решение системы (140) ищется в виде

$$\left.\begin{aligned} y&=\varphi_1(x)y_1+\varphi_2(x)y_2+\varphi_3(x)y_3,\\ z&=\varphi_1(x)z_1+\varphi_2(x)z_2+\varphi_3(x)z_3\\ u&=\varphi_1(x)u_1+\varphi_2(x)u_2+\varphi_3(x)u_3\end{aligned}\right\}$$

После подстановки в (140) получается (проверьте!) система

$$\left.\begin{aligned} \varphi_1'y_1+\varphi_2'y_2+\varphi_3'y_3&=f_1(x),\\ \varphi_1'z_1+\varphi_2'z_2+\varphi_3'z_3&=f_2(x),\\ \varphi_1'u_1+\varphi_2'u_2+\varphi_3'u_3&=f_3(x),\end{aligned}\right\}$$

из которой алгебраическим способом находим φ_1', φ_2', φ_3', а затем, интегрируя, находим φ_1, φ_2, φ_3

Особое значение имеют *линейные однородные системы с постоянными коэффициентами*. Рассмотрим, например, систему

$$\left.\begin{aligned} y'&=a_1y+b_1z+c_1u,\\ z'&=a_2y+b_2z+c_2u,\\ u'&=a_3y+b_3z+c_2u,\end{aligned}\right\}\tag{141}$$

в которой все коэффициенты $a_1, b_1, \ldots, c_3$ постоянны. Она решается подобно п. 17. Именно, ненулевые частные решения ищут в виде

$$y=\lambda e^{px},\quad z=\mu e^{px},\quad u=\nu e^{px},\tag{142}$$

где λ, μ, ν, p — пока неизвестные постоянные. Подстановка в (141) дает после сокращения на e^{px} и переноса всех членов в одну

сторону:

$$\left.\begin{aligned}(a_1-p)\lambda+b_1\mu+c_1\nu=0,\\ a_2\lambda+(b_2-p)\mu+c_2\nu=0,\\ a_3\lambda+b_3\mu+(c_3-p)\nu=0.\end{aligned}\right\}\qquad(143)$$

Эти равенства можно рассматривать как систему трех алгебраических однородных уравнений первой степени с тремя неизвестными λ, μ, ν. Чтобы она имела ненулевое решение, а только такое решение в силу (142) нас и интересует, необходимо и достаточно (см. конец п. VI.6), чтобы определитель системы был равен нулю:

$$\begin{vmatrix} a_1-p & b_1 & c_1\\ a_2 & b_2-p & c_2\\ a_3 & b_3 & c_3-p\end{vmatrix}=0. \qquad (144)$$

Это — *характеристическое уравнение* для системы (141), из которого мы находим возможные значения p.

Так как уравнение (144) имеет третью степень относительно p (почему?), то оно имеет три корня, p_1, p_2, p_3. Если все эти корни простые, то можно любой из них подставить в систему (143), найти какое-либо ненулевое решение λ, μ, ν и по формуле (142) получить соответствующее решение $y(x)$, $z(x)$, $u(x)$. Из построенных таким образом трех частных решений (при $p=p_1$, p_2 и p_3) согласно формуле (139) получаем общее решение системы (141):

$$\left.\begin{aligned}y=C_1\lambda_1e^{p_1x}+C_2\lambda_2e^{p_2x}+C_3\lambda_3e^{p_3x},\\ z=C_1\mu_1e^{p_1x}+C_2\mu_2e^{p_2x}+C_3\mu_3e^{p_3x},\\ u=C_1\nu_1e^{p_1x}+C_2\nu_2e^{p_2x}+C_3\nu_3e^{p_3x},\end{aligned}\right\}\qquad(145)$$

где C_1, C_2, C_3 — произвольные постоянные.

Если уравнение (144) имеет мнимые корни, то решение можно либо оставить в форме (145), либо записать в вещественной форме подобно п. 17 (случай 2).

Случай, когда уравнение (144) имеет кратные корни, более сложный, и мы не будем его здесь рассматривать. В конкретных примерах можно поступать так. Если, например, корень p_1 двойной, то частные решения системы (141) надо при $p=p_1$, взамен (142), искать в форме

$$y=(\lambda x+\overline{\lambda})\,e^{p_1x},\quad z=(\mu x+\overline{\mu})\,e^{p_1x},\quad u=(\nu x+\overline{\nu})\,e^{p_1x}; \qquad (146)$$

после подстановки в (141) приравнять коэффициенты при одинаковых степенях x и из полученной системы уравнений найти *два* независимых варианта значений λ, $\overline{\lambda}$, ..., $\overline{\nu}$ и тем самым два независимых решения вида (146). Если корень характеристического уравнения имеет высшую кратность, то соответственно усложнится и форма частного решения системы (141).

В теории систем линейных дифференциальных уравнений широко применяется матричная запись (см. п. XI.1—2). Для этого систему (138) обычно

переписывают в виде

$$\left.\begin{aligned} y_1' &= a_{11}(x)\,y_1 + a_{12}(x)\,y_2 + a_{13}(x)\,y_3, \\ y_2' &= a_{21}(x)\,y_1 + a_{22}(x)\,y_2 + a_{23}(x)\,y_3, \\ y_3' &= a_{31}(x)\,y_1 + a_{32}(x)\,y_2 + a_{33}(x)\,y_3 \end{aligned}\right\} \qquad (147)$$

и вводят *матрицу коэффициентов* и *векторное решение* (т. е. столбцевую матрицу)

$$\mathbf{A}(x) = \begin{pmatrix} a_{11}(x) & a_{12}(x) & a_{13}(x) \\ a_{21}(x) & a_{22}(x) & a_{23}(x) \\ a_{31}(x) & a_{32}(x) & a_{33}(x) \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Заметим для дальнейшего, что если дана некоторая матрица $\mathbf{B}(x) = (b_{ij}(x))$, то из правил линейных действий с матрицами (см. начало п. XI.2) следует, что $\dfrac{\Delta \mathbf{B}}{\Delta x} = \left(\dfrac{\Delta b_{ij}}{\Delta x}\right)$ и тем самым $\mathbf{B}'(x) = (b'_{ij}(x))$, т. е. *чтобы продифференцировать матрицу, надо продифференцировать все ее элементы.* При этом легко проверить, что основные правила дифференцирования, такие как формулы для производной суммы или произведения и т. п., остаются в силе. Отсюда

$$\mathbf{y}' = \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix}$$

и потому подобно (XI.9) систему (147) можно записать в *матричной форме*

$$\mathbf{y}' = \mathbf{A}(x)\,\mathbf{y}, \qquad (148)$$

а линейную неоднородную систему (140) — в аналогичной форме

$$\mathbf{y}' = \mathbf{A}(x)\,\mathbf{y} + \mathbf{f}(x)$$

Решение (139) можно переписать в векторном виде

$$\mathbf{y} = C_1\mathbf{y}^1 + C_2\mathbf{y}^2 + C_3\mathbf{y}^3,$$

где индексы сверху означают номера линейно независимых частных векторных решений уравнения (148). Система (141) приобретает вид

$$\mathbf{y}' = \mathbf{A}\mathbf{y}, \qquad (149)$$

а решение (142) — вид

$$\mathbf{y} = e^{px}\boldsymbol{\alpha}, \qquad (150)$$

где

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

— некоторый постоянный вектор. Подставляя (150) в (149), получаем

$$pe^{px}\boldsymbol{\alpha} = \mathbf{A}e^{px}\boldsymbol{\alpha}, \quad \text{т. е} \quad \mathbf{A}\boldsymbol{\alpha} = p\boldsymbol{\alpha}.$$

Мы видим (ср. п. XI.4), что $\boldsymbol{\alpha}$ и p должны быть собственным вектором и соответствующим собственным значением матрицы $\mathbf{A}$; последнее, как было показано в п. XI.4, находится из уравнения

$$\det(\mathbf{A} - p\mathbf{I}) = 0,$$

которое есть не что иное, как уравнение (144).

При решении систем вида (141) с постоянными коэффициентами, а также соответствующих неоднородных систем можно применять операторный метод (п. 20). Для этого записываем $y' = Dy$, $z' = Dz$, $u' = Du$, затем решаем

полученную систему уравнений как алгебраическую относительно y, z, u по методам п. VI.4, однако решение доводим лишь до формул вида (VI.13), т е. вида $P(D)y=f(x)$, после чего применяем методы п 20. По существу, это сводится к указанию правила перехода от системы уравнений первого порядка к одному уравнению высшего порядка.

22. Приложение к выяснению устойчивости по Ляпунову состояния равновесия. Понятие устойчивости как способности того или иного объекта состояния или процесса сопротивляться не учитываемым заранее внешним воздействиям появилось еще в античной науке и сейчас занимает одно из центральных мест в физике и технике. Существуют различные конкретные реализации этого общего понятия в зависимости от типа рассматриваемого объекта, характера внешних воздействий и г д. Одна из таких реализаций появлялась у нас в п 16. Сейчас мы рассмотрим понятие *устойчивости по Ляпунову*, одно из наиболее важных, введенное и систематически изученное А. М. Ляпуновым в 1892 г.

Пусть состояние некоторого объекта описывается конечным числом параметров, для определенности тремя параметрами, x, y, z, так что изменение этого объекта во времени задается тремя функциями $x=x(t)$, $y=y(t)$, $z=z(t)$ (t—время). Пусть закон этого изменения имеет вид системы дифференциальных уравнений

$$\left.\begin{aligned}\frac{dx}{dt}&=P(x,\ y,\ z),\\ \frac{dy}{dt}&=Q(x,\ y,\ z),\\ \frac{dz}{dt}&=R(x,\ y,\ z)\end{aligned}\right\} \tag{151}$$

с заданными правыми частями, не содержащими явно независимой переменной t. Последнее условие означает, что дифференциальный закон развития процесса не меняется с течением времени.

Пусть состояние *равновесия* рассматриваемого объекта, когда он не меняется с течением времени, описывается постоянными значениями $x=x_0$, $y=y_0$, $z=z_0$; тогда эта система постоянных, рассматриваемых как функции времени, также должна удовлетворять системе (151). Из непосредственной подстановки в (151) следует, что для этого необходимо и достаточно, чтобы одновременно

$$P(x_0,\ y_0,\ z_0)=0, \quad Q(x_0,\ y_0,\ z_0)=0, \quad R(x_0,\ y_0,\ z_0)=0. \tag{152}$$

Пусть в некоторый момент t_0 объект под влиянием каких то причин вышел из состояния равновесия, т. е. параметры x, y, z стали равными $x=x_0+\Delta x_0$, $y=y_0+\Delta y_0$, $z=z_0+\Delta z_0$ Чтобы выяснить дальнейшее изменение рассматриваемого объекта, надо решить систему уравнений (151) при начальных условиях

$$x(t_0)=x_0+\Delta x_0, \quad y(t_0)=y_0+\Delta y_0, \quad z(t_0)=z_0+\Delta z_0. \tag{153}$$

Исследуемое состояние равновесия называется устойчивым по Ляпунову, если после бесконечно малого выхода из этого состояния объект продолжает оставаться в бесконечной близости от него на протяжении всего дальнейшего времени. Другими словами, при бесконечно малых Δx_0, Δy_0, Δz_0 для решения системы (151) при начальных условиях (153) разности

$$\Delta x=x(t)-x_0, \quad \Delta y=y(t)-y_0, \quad \Delta z=z(t)-z_0$$

должны быть бесконечно малыми на всем интервале времени $t_0\leqslant t<\infty$.

На первый взгляд может показаться странным рассмотрение бесконечно малых отклонений параметров и бесконечно большого промежутка времени,

так как на практике все эти величины конечны. Однако полезно вспомнить различие практической и математической бесконечностей (п. III.1). Практической бесконечно малой является просто малая в масштабах рассматриваемого процесса реальная величина, а практически бесконечным промежутком времени является время *переходного процесса*, т. е. перехода от исследуемого состояния к состоянию иного типа (например, от одного состояния равновесия к другому или от состояния равновесия к разрушению объекта и т. п). Таким образом, реально устойчивость по Ляпунову означает, что малый выход из состояния равновесия практически не нарушает этого состояния

Для выяснения того, будет ли иметь место устойчивость, подставим в систему (151) $x=x_0+\Delta x$, $y=y_0+\Delta y$, $z=z_0+\Delta z$, что даст

$$\left.\begin{aligned} \frac{d(\Delta x)}{dt} &= P(x_0+\Delta x,\ y_0+\Delta y,\ z_0+\Delta z)= \\ &= (P'_x)_0\,\Delta x+(P'_y)_0\,\Delta y+(P'_z)_0\,\Delta z+\ldots, \\ \frac{d(\Delta y)}{dt} &= (Q'_x)_0\,\Delta x+(Q'_y)_0\,\Delta y+(Q'_z)_0\,\Delta z+\ldots, \\ \frac{d(\Delta z)}{dt} &= (R'_x)_0\,\Delta x+(R'_y)_0\,\Delta y+(R'_z)_0\,\Delta z+\ldots, \end{aligned}\right\} \tag{154}$$

где обозначено $(P'_x)_0=P'_x(x_0,\ y_0,\ z_0)$ и т. п. Здесь при преобразовании правых частей мы воспользовались формулой Тейлора (XII 17) и формулами (152); многоточиями обозначены члены выше первого порядка малости

Так как при выяснении устойчивости рассматриваются лишь малые Δx, Δy, Δz, то в правых частях системы (154) основную роль играют выписанные, линейные члены. Поэтому заменим систему (154) на *укороченную систему* (*систему первого приближения*), отбросив члены высшего порядка малости:

$$\left.\begin{aligned} \frac{d(\Delta x)}{dt} &= (P'_x)_0\,\Delta x+(P'_y)_0\,\Delta y+(P'_z)_0\,\Delta z, \\ \frac{d(\Delta y)}{dt} &= (Q'_x)_0\,\Delta x+(Q'_y)_0\,\Delta y+(Q'_z)_0\,\Delta z, \\ \frac{d(\Delta z)}{dt} &= (R'_x)_0\,\Delta x+(R'_y)_0\,\Delta y+(R'_z)_0\,\Delta z. \end{aligned}\right\} \tag{155}$$

Система (155) — это линейная система с постоянными коэффициентами, которая решается по методу п 21. Согласно формуле (145) (где, правда, были применены иные обозначения) решение системы (155) получается как комбинация функций вида e^{pt}, где p удовлетворяет характеристическому уравнению

$$\begin{vmatrix} (P'_x)_0-p & (P'_y)_0 & (P'_z)_0 \\ (Q'_x)_0 & (Q'_y)_0-p & (Q'_z)_0 \\ (R'_x)_0 & (R'_y)_0 & (R'_z)_0-p \end{vmatrix}=0. \tag{156}$$

При этом малым Δx_0, Δy_0, Δz_0 отвечают малые значения произвольных постоянных C_1, C_2, C_3, и поэтому все дело в поведении функции e^{pt} при возрастании t Так как при $p=r+is$ (случай $s=0$ не исключен) будет $|e^{pt}|=e^{rt}$ (см. формулу (VIII 13)), то

$$|e^{pt}|_{t\to\infty}\longrightarrow 0 \text{ (при } r<0), \quad |e^{pt}|_{t\to\infty}\longrightarrow\infty \text{ (при } r>0), \tag{157}$$

и мы приходим к следующим выводам. *Если все корни характеристического*

уравнения (156) *имеют отрицательную вещественную часть* (в частности, они могут быть вещественными отрицательными), *то рассматриваемое состояние равновесия* x_0, y_0, z_0 *устойчиво по Ляпунову*. Кроме того, в этом случае при малых Δx_0, Δy_0, Δz_0, будет $x(t) \to x_0$, $y(t) \to y_0$, $z(t) \to z_0$ при $t \to \infty$; такая устойчивость называется *асимптотической*. *Если же среди корней уравнения* (156) *имеется по крайней мере один с положительной вещественной частью, то рассматриваемое состояние равновесия неустойчиво по Ляпунову*

Эти результаты мы вывели для системы (155), но согласно сказанному выше те же утверждения справедливы для полной системы (154). Отметим, что наличие кратных корней у уравнения (156) не нарушает наших утверждений, несмотря на то, что при этом в решении могут появиться степени t в качестве множителей, так как экспонента (157) при $r < 0$ стремится к нулю быстрее любой степени t.

Оба полученных вывода не охватывают случая, когда среди корней уравнения (156) нет корней с положительной вещественной частью, но имеется по крайней мере один с нулевой вещественной частью. Тогда в общем решении системы (155) появляются функции вида

$$e^{ist} = \cos st + i \sin st, \qquad |e^{ist}| = 1,$$

т. е. получается, будто бы рассматриваемый объект колеблется или остается неподвижным около состояния равновесия, не стремясь к нему. Но тогда из-за неограниченности времени начинают влиять отброшенные члены высшего порядка малости, которые могут нарушить устойчивость. Итак, *в рассматриваемом особом случае по корням уравнения* (156) *нельзя заключить об устойчивости или неустойчивости состояния равновесия*; чтобы это сделать, надо привлечь какие-либо дополнительные соображения, например привлечь дальнейшие члены разложений (154).

Полученные результаты приобретают особенно простой вид для случая, когда изменение объекта описывается одной функцией $x(t)$, удовлетворяющей дифференциальному уравнению

$$\frac{dx}{dt} = f(x). \tag{158}$$

Мы получаем, что если $f(x_0) = 0$, $f'(x_0) < 0$, то значению $x = x_0$ отвечает устойчивое состояние равновесия, а если $f(x_0) = 0$, $f'(x_0) > 0$, то это состояние неустойчивое. (Получите этот вывод, исходя из расположения изоклин для уравнения (158) на плоскости t, x.)

По теории устойчивости имеется большое количество руководств; укажем, в частности, на наиболее полный курс: И. Г. Малкин, Теория устойчивости движения, изд. 2-е, «Наука», 1966.

§ 7. Приближенное и численное решение уравнений

Часто бывает невозможно осуществить точное интегрирование дифференциального уравнения в квадратурах. Тогда для построения решения приходится применять иные методы. В п. 3 был описан простейший графический метод решения уравнений первого порядка. Здесь мы укажем некоторые *методы построения приближенных формул* для решения, аналогичные описанным в § V.1 методам решения конечных уравнений, а также *методы численного решения*, в которых искомое частное решение строится в табличном виде. Мы будем для простоты рассматривать уравнения первого порядка, но те

же методы естественно переносятся на уравнения любого порядка и на системы уравнений.

23. Метод итераций. Рассмотрим дифференциальное уравнение первого порядка с заданным начальным условием

$$y' = f(x,\ y), \quad y(x_0) = y_0. \tag{159}$$

Если взять интегралы от обеих частей уравнения, получим

$$\int_{x_0}^{x} y'\, dx = y - y_0 = \int_{x_0}^{x} f(x,\ y)\, dx = \int_{x_0}^{x} f(x,\ y(x))\, dx.$$

Изменяя обозначение переменной интегрирования, напишем

$$y(x) = y_0 + \int_{x_0}^{x} f(s,\ y(s))\, ds. \tag{160}$$

Уравнение (160) равносильно сразу обоим равенствам (159), так как после его дифференцирования получится первое равенство, а после подстановки $x = x_0$ — второе. Уравнение (160) является *интегральным уравнением*, так как в нем неизвестная функция стоит под знаком интеграла.

Вид уравнения (160) удобен для применения метода итераций (ср. с уравнением (V. 9), хотя сейчас неизвестной является не число, а функция). Выбрав некоторую функцию $y_0(x)$ в качестве нулевого приближения (желательно, чтобы она была по возможности ближе к искомому решению; если о последнем ничего не известно, то можно положить хотя бы $y_0(x) \equiv y_0$), находим первое приближение по формуле

$$y_1(x) = y_0 + \int_{x_0}^{x} f(s,\ y_0(s))\, ds.$$

Подставляя результат в правую часть (160), находим второе приближение и т. д.; вообще

$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f(s,\ y_n(s))\, ds \quad (n = 0,\ 1,\ 2,\ \ldots). \tag{161}$$

Подобно п. V.3, если процесс итераций сходится, т. е. если последовательные приближения стремятся с ростом n к некоторой предельной функции, то она удовлетворяет уравнению (160); для проверки этого надо в равенстве (161) перейти к пределу при $n \to \infty$.

Замечательно, что *метод итераций для уравнения* (160), *как правило, сходится для всех x, достаточно близких к x_0*; так будет, во всяком случае, если выполнены условия теоремы Коши из п. 3. Это связано с тем, что при вычислении последующих приближений

надо интегрировать предыдущие, а при последовательном интегрировании функции в целом «сглаживаются» и всякие неправильности, происходящие из-за выбора нулевого приближения, погрешностей округления и т. п., постепенно устраняются. В отличие от этого при последовательном дифференцировании функции, как правило, ухудшаются, первоначальные неправильности разрастаются и поэтому итерационный метод, основанный на последовательном дифференцировании, не дал бы сходимости.

Рис. 298.

Различие между интегрированием и дифференцированием показано на рис. 298. Нарисованный «горбик», добавленный к какой-либо функции, значительно портит ее производную, не говоря уже о последующих производных (какой вид имеет вторая производная $y=f''(x)$?) и почти не сказывается на интеграле.

Рассмотрим, например, частный вид уравнения Риккати (см. конец п. 4)

$$y'=x^2+y^2$$

при начальном условии $y(0)=0$. После интегрирования получим

$$y(x)=\frac{x^3}{3}+\int_0^x y^2(s)\,ds.$$

Выберем в качестве нулевого приближения для искомого решения, о котором мы пока ничего не знаем, нулевую функцию $y_0(x)\equiv 0$, так как она удовлетворяет хотя бы начальному условию. Тогда получим (проверьте!)

$$y_1(x)=\frac{x^3}{3},\quad y_2(x)=\frac{x^3}{3}+\int_0^x \left(\frac{s^3}{3}\right)^2 dx=\frac{x^3}{3}+\frac{x^7}{63},$$

$$y_3(x)=\frac{x^3}{3}+\frac{x^7}{63}+\frac{2x^{11}}{2079}+\frac{x^{15}}{59535}$$

и т. д. Видно, что при небольших x, например при $|x|<1$, процесс хорошо сходится; так, с точностью до 0,001 при $|x|<1$ можно положить $y=\frac{x^3}{3}+\frac{x^7}{63}$, а при $|x|<\frac{1}{2}$ даже просто $y=\frac{x^3}{3}$.

Как обычно на практике, вопрос о том, на каком приближении нужно остановиться, решается с помощью сравнения последующих приближений с предыдущими.

24. Применение ряда Тейлора. Из уравнения и начального условия (159) можно с помощью дифференцирования найти значения $y'(x_0)$, $y''(x_0)$ и т. д., после чего составить разложение решения в степенной ряд Тейлора (п. IV.16). Необходимое количество членов определяется с помощью их последовательного вычисления и сравнения с выбранной степенью точности.

Рассмотрим, например, задачу

$$y' = x^2 + y^2, \quad y(0) = 1.$$

Подстановкой в правую часть уравнения находим, что

$$y'(0) = 0^2 + 1^2 = 1.$$

Если продифференцировать обе части уравнения, получим $y'' = 2x + 2yy'$ и, подставив $x = 0$, найдем

$$y''(0) = 2\cdot 0 + 2\cdot 1\cdot 1 = 2.$$

Аналогично находим

$$y''' = 2 + 2y'^2 + 2yy''; \quad y'''(0) = 8; \quad y^{IV} = 6y'y'' + 2yy''', \quad y^{IV}(0) = 28$$

и т. д. Подставляя это в формулу Маклорена (IV.54), получим

$$y = y(0) + \frac{y'(0)}{1!}x + \frac{y''(0)}{2!}x^2 + \ldots = 1 + x + x^2 + \frac{4}{3}x^3 + \frac{7}{6}x^4 + \ldots$$

Этой формулой можно пользоваться для небольших $|x|$.

25. Применение степенных рядов с неопределенными коэффициентами. Этот метод тесно связан с п. 24 и состоит в том, что решение уравнения ищется в форме ряда с неизвестными коэффициентами

$$y = a + b(x - x_0) + c(x - x_0)^2 + d(x - x_0)^3 + \ldots, \qquad (162)$$

которые находятся с помощью подстановки в уравнение и последующего приравнивания коэффициентов при одинаковых степенях x (и применения начального условия, если оно задано).

Рассмотрим, например, уравнение второго порядка

$$y'' + xy = 0. \qquad (163)$$

Будем искать решение разложенным по степеням x:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots \qquad (164)$$

После дифференцирования и подстановки в уравнение получим

$$(1\cdot 2a_2 + 2\cdot 3a_3x + 3\cdot 4a_4x^2 + \ldots) + x(a_0 + a_1x + a_2x^2 + \ldots) = 0.$$

Приравнивание коэффициентов при одинаковых степенях x дает

$$1\cdot 2a_2 = 0; \quad 2\cdot 3a_3 + a_0 = 0; \quad 3\cdot 4a_4 + a_1 = 0;$$
$$4\cdot 5a_5 + a_2 = 0; \quad 5\cdot 6a_6 + a_3 = 0, \ \ldots,$$

откуда последовательно находим

$$a_2 = 0; \quad a_3 = -\frac{a_0}{2\cdot 3}; \quad a_4 = -\frac{a_1}{3\cdot 4}; \quad a_5 = -\frac{a_2}{4\cdot 5} = 0;$$
$$a_6 = -\frac{a_3}{5\cdot 6} = \frac{a_0}{2\cdot 3\cdot 5\cdot 6}; \quad a_7 = -\frac{a_4}{6\cdot 7} = \frac{a_0}{3\cdot 4\cdot 6\cdot 7}; \quad a_8 = -\frac{a_5}{7\cdot 8} = 0;$$
$$a_9 = -\frac{a_6}{8\cdot 9} = -\frac{a_0}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} \text{ и т. д.}$$

Подстановка этих результатов в формулу (164) дает общее решение уравнения (163):

$$y = a_0 + a_1 x - \frac{a_0}{2\cdot 3} x^3 - \frac{a_1}{3\cdot 4} x^4 + \frac{a_0}{2\cdot 3\cdot 5\cdot 6} x^6 + \\ + \frac{a_1}{3\cdot 4\cdot 6\cdot 7} x^7 - \frac{a_0}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} x^9 - \ldots = \\ = a_0 \left(1 - \frac{x^3}{2\cdot 3} + \frac{x^6}{2\cdot 3\cdot 5\cdot 6} - \frac{x^9}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} + \ldots\right) + \\ + a_1 \left(x - \frac{x^4}{3\cdot 4} + \frac{x^7}{3\cdot 4\cdot 6\cdot 7} - \frac{x^{10}}{3\cdot 4\cdot 6\cdot 7\cdot 9\cdot 10} + \ldots\right).$$

Константы a_0 и a_1 остаются в качестве произвольных постоянных (в свойстве 5 п. 14 они были обозначены через C_1 и C_2). Ряды, стоящие в скобках, представляют два линейно независимых частных решения уравнения (163).

Описанный прием всегда применим, в частности, к линейным уравнениям

$$a_0(x)\, y^{(n)} + a_1(x)\, y^{(n-1)} + \ldots + a_n(x)\, y = f(x), \qquad (165)$$

если все функции $a_0(x)$, $a_1(x)$, ..., $f(x)$ представляют собой многочлены относительно x или, в более общем случае, суммы рядов по степеням $x - x_0$, причем $a_0(x_0) \neq 0$.

Если $a_0(x_0) = 0$, то значение x_0 называется *особой точкой* для уравнения (165); тогда найти решение в форме (162) возможно не всегда. В этом случае иногда удается найти решение в форме

$$y = (x - x_0)^\rho \,[a + b(x - x_0) + c(x - x_0)^2 + d(x - x_0)^3 + \ldots], \qquad (166)$$

где постоянная ρ также подбирается. При этом можно считать, что $a \neq 0$, так как в противном случае можно вынести за скобку некоторую степень $x - x_0$, так что дело сведется к изменению ρ.

26. Функции Бесселя. Рассмотрим важный пример уравнения Бесселя

$$x^2 y'' + x y' + (x^2 - p^2)\, y = 0 \qquad (p = \mathrm{const} \geqslant 0;\ 0 < x < \infty). \qquad (167)$$

Решения этого уравнения называются *функциями Бесселя*, хотя они применялись Эйлером с 1766 г., т. е. до рождения Бесселя. Эти функции называются также *цилиндрическими* функциями, так как они широко применяются при решении уравнений математической физики в круглом цилиндре.

Значение $x = 0$ является для уравнения (167) особой точкой, поэтому в силу формулы (166), в которой надо положить $x_0 = 0$, его решение можно искать в виде

$$y = a x^\rho + b x^{\rho+1} + c x^{\rho+2} + \ldots \qquad (168)$$

Дифференцирование и подстановка в уравнение (167) дают

$$x^2 [a\rho(\rho - 1) x^{\rho-2} + b(\rho + 1)\rho x^{\rho-1} + c(\rho + 2)(\rho + 1) x^\rho + \ldots] + \\ + x [a\rho x^{\rho-1} + b(\rho + 1) x^\rho + c(\rho + 2) x^{\rho+1} + \ldots] + \\ + x^2 (a x^\rho + b x^{\rho+1} + c x^{\rho+2} + \ldots) - p^2 (a x^\rho + b x^{\rho+1} + c x^{\rho+2} + \ldots) = 0.$$

После приравнивания нулю коэффициентов при одинаковых степенях x получим цепочки равенств:

$$a\rho(\rho - 1) + a\rho - ap^2 = 0, \text{ т. е. } a(\rho^2 - p^2) = 0;$$
$$b(\rho + 1)\rho + b(\rho + 1) - bp^2 = 0, \text{ т. е. } b(\rho^2 + 2\rho + 1 - p^2) = 0;$$

$$c(\rho+2)(\rho+1)+c(\rho+2)+a-cp^2=0, \quad \text{т. е.} \quad c(\rho^2+4\rho+4-p^2)+a=0;$$
$$d(\rho+3)(\rho+2)+d(\rho+3)+b-dp^2=0, \quad \text{т. е.} \quad d(\rho^2+6\rho+9-p^2)+b=0;$$
$$e(\rho+4)(\rho+3)+e(\rho+4)+c-ep^2=0, \quad \text{т. е.} \quad e(\rho^2+8\rho+16-p^2)+c=0$$

и т. д. Из первого равенства, поскольку $a \neq 0$, мы видим, что $\rho^2=p^2$, т. е. $\rho=\pm p$. Подставляя этот результат в остальные равенства, получим последовательно

$$b=0, \quad c=\frac{-a}{4\rho+4}=-\frac{a}{2^2(\rho+1)}, \quad d=0,$$
$$e=-\frac{c}{8\rho+16}=\frac{a}{2^4\cdot 2(\rho+1)(\rho+2)}, \quad f=0,$$
$$g=-\frac{a}{2^6\cdot 2\cdot 3(\rho+1)(\rho+2)(\rho+3)}, \quad i=0,$$
$$j=\frac{a}{2^8\cdot 2\cdot 3\cdot 4(\rho+1)(\rho+2)(\rho+3)(\rho+4)} \quad \text{и т. д.}$$

Отсюда в силу формулы (168) находим решение

$$y=ax^\rho-\frac{a}{2^2(\rho+1)}x^{\rho+2}+\frac{a}{2^4\cdot 2!(\rho+1)(\rho+2)}x^{\rho+4}-$$
$$-\frac{a}{2^6\cdot 3!(\rho+1)(\rho+2)(\rho+3)}x^{\rho+6}+\ldots, \quad (169)$$

где a — произвольная постоянная. Удобно выбрать $a=\frac{1}{2^\rho\Gamma(\rho+1)}$ (п. XIV.17). Если учесть, что в силу формулы (XIV.66)

$$\Gamma(\rho+1)(\rho+1)=\Gamma(\rho+2); \quad \Gamma(\rho+1)(\rho+1)(\rho+2)=$$
$$=\Gamma(\rho+2)(\rho+2)=\Gamma(\rho+3) \text{ и т. д.,}$$

то формула (169) при таком a даст

$$y=\frac{1}{\Gamma(\rho+1)}\left(\frac{x}{2}\right)^\rho-\frac{1}{1!\Gamma(\rho+2)}\left(\frac{x}{2}\right)^{\rho+2}+\frac{1}{2!\Gamma(\rho+3)}\left(\frac{x}{2}\right)^{\rho+4}-$$
$$-\frac{1}{3!\Gamma(\rho+4)}\left(\frac{x}{2}\right)^{\rho+6}+\ldots=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!\Gamma(\rho+n+1)}\left(\frac{x}{2}\right)^{\rho+2n}. \quad (170)$$

Эта сумма называется *функцией Бесселя 1-го рода порядка* ρ и обозначается через $J_\rho(x)$. Так как $\rho=\pm p$, то общее решение уравнения (167) (п. 14, свойство 5) можно записать в виде

$$y=C_1J_p(x)+C_2J_{-p}(x). \quad (171)$$

Решение (171) не годится при целом $p=0, 1, 2, 3, \ldots$ Действительно, для таких p при $\rho=-p$ будет

$$\Gamma(-p+1)=\Gamma(-p+2)=\ldots=\Gamma(-p+p)=\pm\infty,$$

и потому формула (172) даст

$$J_{-p}(x)=\sum_{n=p}^{\infty}\frac{(-1)^n}{n!(-p+n)!}\left(\frac{x}{2}\right)^{-p+2n}=|n-p=n'|=$$
$$=\sum_{n'=0}^{\infty}\frac{(-1)^p(-1)^{n'}}{(p+n')!(n')!}\left(\frac{x}{2}\right)^{p+2n'}=(-1)^pJ_p(x) \quad (p=0, 1, 2, \ldots).$$

Значит, в этом случае решения $J_p(x)$ и $J_{-p}(x)$ линейно зависимые (п. 14), а потому формула (171) не дает общего решения.

Чтобы получить общее решение уравнения (167), пригодное для всех p, поступают аналогично случаю 3 п. 17. Именно, сначала считают, что p не целое, и образуют функцию

$$Y_p(x) = \operatorname{ctg} p\pi J_p(x) - \frac{1}{\sin p\pi} J_{-p}(x) = \frac{\cos p\pi J_p(x) - J_{-p}(x)}{\sin p\pi}.$$

Как линейная комбинация решений она также является решением уравнения (167) и называется *функцией Бесселя 2 го рода порядка p*; она иногда обозначается также через $N_p(x)$. Если p становится целым, то в правой части получается неопределенность (почему?). Ее можно раскрыть по правилу Лопиталя, чего мы здесь делать не будем. Отметим только, что в итоге получится сумма, для которой при $x \to 0$ старшим членом будет

$$-\frac{(p-1)!\,2^p}{\pi x^p} \quad (p=1, 2, 3, \ldots); \qquad \frac{2}{\pi} \ln x \quad (p=0).$$

Итак, формула

$$y = C_1 J_p(x) + C_2 Y_p(x) \tag{172}$$

дает общее решение уравнения (167) для всех $p \geqslant 0$, нецелых или целых, при $0 < x < \infty$. При этом $J_p(+0)$ конечно, тогда как $Y_p(+0) = -\infty$. Поэтому если $y(+0)$ по условиям задачи должно быть конечным, то в правой части формулы (172) надо оставить лишь первое слагаемое.

В настоящее время функции Бесселя подробно изучены и затабулированы. Наибольшее значение для приложений имеют

$$\left.\begin{aligned} J_0(x) &= 1 - \frac{x^2}{1!^2 2^2} + \frac{x^4}{2!^2 2^4} - \frac{x^6}{3!^2 2^6} + \ldots, \\ J_1(x) &= \frac{x}{2} - \frac{x^3}{1!2!2^3} + \frac{x^5}{2!3!2^5} - \frac{x^7}{3!4!2^7} + \ldots \end{aligned}\right\} \tag{173}$$

Примерные графики этих функций показаны на рис. 299. Эти функции, как и все функции Бесселя 1-го и 2-го рода, при возрастании x бесконечное

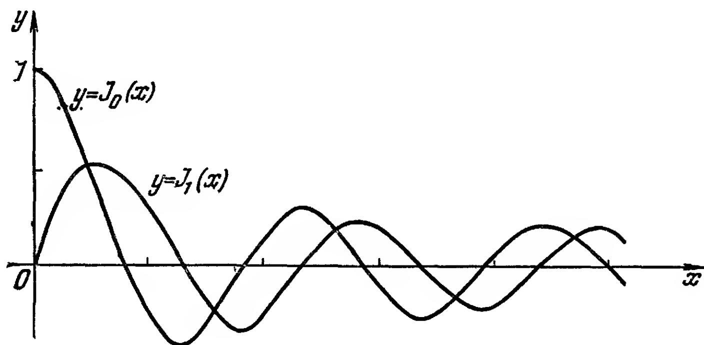


Рис. 299.

число раз меняют знак и стремятся к нулю. Из формул (173) легко вывести, что $J_0'(x) = -J_1(x)$ (проверьте!). Имеются и другие соотношения между функциями Бесселя. Все эти свойства можно найти, например, в книге [29].

27. Метод малого параметра. Этот метод, описанный в п. V.5, применяется и при решении дифференциальных уравнений. Приведем простые примеры

Задача

$$y' = \frac{x}{1+0,1xy}, \qquad y(0)=0 \tag{174}$$

не содержит параметров. Однако можно рассмотреть более общую задачу

$$y' = \frac{x}{1+\alpha xy}, \qquad y(0)=0, \tag{175}$$

из которой (174) получается при $\alpha=0,1$. Задача (175) легко решается при $\alpha=0$: тогда получается $y=\frac{x^2}{2}$. Поэтому ищем решение задачи разложенным в ряд по степеням α, т. е.

$$y=\frac{x^2}{2}+\alpha u+\alpha^2 v+\alpha^3 w+\ldots, \tag{176}$$

где $u=u(x)$, $v=v(x)$ и т. д. — пока неизвестные функции x.

Подстановка (176) в (175) дает после умножения на знаменатель

$$(x+\alpha u'+\alpha^2 v'+\alpha^3 w'+\ldots)\left(1+\frac{\alpha}{2}x^3+\alpha^2 xu+\alpha^3 xv+\ldots\right)=x; \tag{177}$$

$$\alpha u(0)+\alpha^2 v(0)+\ldots=0, \quad \text{т. е.} \quad u(0)=0, \quad v(0)=0, \quad w(0)=0, \ \ldots \tag{178}$$

Раскрывая скобки в (177) и приравнивая нулю коэффициенты при степенях α, получим последовательно

$$u'+\frac{1}{2}x^4=0, \quad v'+\frac{x^3}{2}u'+x^2u=0, \quad w'+\frac{x^3}{2}v'+xuu'+x^2v=0 \text{ и т. д.,}$$

откуда с учетом равенств (178) найдем (проверьте!)

$$u=-\frac{x^5}{10}, \quad v=\frac{7}{160}x^8, \quad w=\frac{71}{1760}x^{11} \text{ и т. д.}$$

Поэтому формула (176) дает

$$y=\frac{x^2}{2}-\frac{\alpha}{10}x^5+\frac{7\alpha^2}{160}x^8-\frac{71\alpha^3}{1760}x^{11}+\ldots$$

В частности, для уравнения (174) получим

$$y=\frac{x^2}{2}-\frac{x^5}{100}+\frac{7x^8}{16\,000}-\frac{71x^{11}}{1\,760\,000}+\ldots$$

Этот ряд прекрасно сходится при $|x|<1$ и неплохо при $1<|x|<2$.

Рассмотрим в качестве другого примера задачу

$$y'=\sin(xy), \qquad y(0)=\alpha. \tag{179}$$

В отличие от предыдущего примера здесь параметр входит в начальное условие. При $\alpha=0$ задача (179) имеет, очевидно, решение $y\equiv 0$. Поэтому при малых $|\alpha|$ ищем решение в форме

$$y=\alpha u+\alpha^2 v+\alpha^3 w+\ldots \qquad (u=u(x),\ v=v(x),\ \ldots). \tag{180}$$

Подстановка значения $x=0$ дает

$$u(0)=1, \quad v(0)=0, \quad w(0)=0, \ \ldots \tag{181}$$

С другой стороны, подставив (180) в дифференциальное уравнение (179), получим с учетом степенно́го ряда для синуса (формула (IV.57)):

$$\alpha u' + \alpha^2 v' + \alpha^3 w' + \ldots = \\ = \frac{(\alpha x u + \alpha^2 x v + \alpha^3 x w + \ldots)}{1!} - \frac{(\alpha x u + \alpha^2 x v + \alpha^3 x w + \ldots)^3}{3!} + \ldots$$

Приравнивание коэффициентов при одинаковых степенях α дает

$$u' = xu, \quad v' = xv, \quad w' = xw - \frac{x^3 u^3}{3!}, \ldots$$

Интегрируя эти уже *линейные* уравнения с учетом начальных условий (181), найдем (проверьте!)

$$u = e^{\frac{x^2}{2}}, \quad v = 0, \quad w = \frac{1}{12}(1 - x^2)\, e^{\frac{3}{2}x^2} - \frac{1}{2} e^{\frac{x^2}{2}}.$$

Подстановка этих выражений в (180) дает разложение искомого решения, пригодное для небольших $|x|$ и $|\alpha|$.

В более сложных случаях при применении метода малого параметра часто бывает полезно найти хотя бы первый содержащий параметр член разложения.

28. Общие замечания о зависимости решения от параметра. В связи с предыдущим пунктом выскажем несколько общих соображений. Часто бывает, что изучаемое дифференциальное уравнение или система таких уравнений содержат один или несколько параметров, которые могут принимать различные постоянные значения. Рассмотрим для простоты уравнение первого порядка

$$\frac{dy}{dx} = f(x, y; \lambda) \qquad (182)$$

(λ — параметр) при определенных начальных условиях $x = x_0$, $y = y_0$.

Будем считать, что точка $(x_0; y_0)$ — неособая (п. 7), т. е. при заданных условиях существует единственное решение уравнения (182). Тогда из геометрического смысла уравнения (182) (п. 3) следует, что если его правая часть зависит от λ непрерывно, то при малом изменении λ поле направлений будет меняться мало, а потому и решение $y(x; \lambda)$ будет зависеть от λ непрерывно. Аналогичный вывод получается, если от λ зависит не только уравнение, но и начальное условие, т. е. если $x_0 = x_0(\lambda)$, $y_0 = y_0(\lambda)$.

Пусть решение $y(x; \lambda)$ уравнения (182) известно при некотором, как говорят, «невозмущенном» значении λ; пусть, далее, значение параметра изменилось и стало равным $\lambda + \Delta\lambda$, где $|\Delta\lambda|$ мало́. Тогда и y изменится и получит приращение $\Delta_\lambda y$, главную линейную часть, т. е. дифференциал которого мы обозначим через δy и назовем *вариацией* решения.

Таким образом, вариация — это частный дифференциал, взятый по параметру; новое название и новое обозначение применяются, чтобы отличить дифференциал по независимой переменной от дифференциала по параметру. В тех случаях, когда малыми высшего порядка можно пренебречь, можно сказать просто, что вариация решения — это бесконечно малое его изменение, полученное за счет изменения параметра. При выбранном значении λ величина δy, как и y, зависит от x, т. е. $\delta y = \delta y(x)$ и прямо пропорциональна $\Delta\lambda$.

Чтобы составить дифференциальное уравнение для δy, надо приравнять дифференциалы обеих частей равенства (182) по λ:

$$\delta \frac{dy}{dx} = \delta(f(x, y; \lambda)) = \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial \lambda}\delta\lambda \qquad (\delta\lambda = \Delta\lambda),$$

т. е.

$$\frac{d(\delta y)}{dx}=f'_y(x, y; \lambda)\,\delta y+f'_\lambda(x, y; \lambda)\,\delta\lambda. \tag{183}$$

При этом в левой части мы переставили знаки d и δ, как знаки дифференциалов по разным переменным (п. IX.15), а в правой части воспользовались формулой для производной сложной функции. Уравнение (183) называется *уравнением в вариациях* для исходного уравнения (182). Так как в правой части взамен y надо подставить «невозмущенное решение» $y(x; \lambda)$, то уравнение (183) линейное и потому легко интегрируется (п. 4). Для уравнений высших порядков и для систем уравнений соответствующие уравнения в вариациях в общем случае не интегрируются в квадратурах, но они *всегда являются линейными.*

Выведем *начальное условие для* δy. В общем случае, когда $x_0=x_0(\lambda)$, $y_0=y_0(\lambda)$, получаем при значении параметра $\lambda+\delta\lambda$, отбрасывая малые высшего порядка, что при $x=x_0(\lambda+\delta\lambda)=x_0+x_0'\delta\lambda$ будет $y=y_0(\lambda+\delta\lambda)=y_0+$ $+y_0'\delta\lambda$, где $x_0'=x_0'(\lambda)$, $y_0'=y_0'(\lambda)$. Отсюда при значении параметра $\lambda+\delta\lambda$ и при $x=x_0$ будет

$$y\,|_{x=x_0}=y\,|_{x=x_0+x_0'\delta\lambda}-\partial_x y\,|_{dx=x_0'\delta\lambda}=y_0+y_0'\delta\lambda-\frac{dy}{dx}x_0'\delta\lambda=y_0+y_0'\delta\lambda-f_0x_0'\delta\lambda,$$

где $f_0=f(x_0, y_0; \lambda)$. Но то же значение y равно $(y\,|_\lambda+\delta y)_{x=x_0}=y_0+(\delta y)_{x=x_0}$. Значит, начальное условие для δy таково:

$$(\delta y)_{x=x_0}=(y_0'-f_0x_0')\,\delta\lambda.$$

В том частном случае, когда x_0 и y_0 не зависят от λ, будет $x_0'=y_0'=0$ и потому начальное условие имеет вид $(\delta y)_{x=x_0}=0$.

Иногда параметр входит в дифференциальное уравнение таким способом, что при некоторых значениях этого параметра уравнение понижает свой порядок, т. е. вырождается. При этом возникают новые обстоятельства, которые мы поясним на примере.

Рассмотрим задачу

$$\lambda y'+y=0, \quad y\,|_{x=0}=1 \tag{184}$$

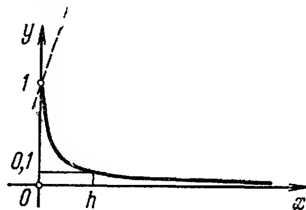


Рис. 300.

с решением $y=e^{-\frac{x}{\lambda}}$. При $\lambda=0$ получается вырождение (почему?). Пусть решение рассматривается при $x\geqslant 0$ и $\lambda\to+0$; это решение показано на рис. 300. Уравнение (184) в пределе переходит в равенство $y=0$, но мы видим, что при малом λ решение близко к нулю не сразу от $x=0$, а только от некоторого $x=h$. Промежуток $0<x<h$, называемый *пограничным слоем*, нужен решению для того, чтобы от единичного начального значения (184) перейти к значению, близкому к нулю.

Ширина пограничного слоя условна, так как теоретически решение нигде не становится точно равным нулю. Если, например, принять за ширину пограничного слоя значение $x=h$, при котором решение уменьшается в 10 раз по сравнению с исходным значением, то для задачи (184) мы получим

$$e^{-h/\lambda}=0{,}1; \qquad h=\ln 10\cdot\lambda,$$

т. е. *ширина пограничного слоя прямо пропорциональна значению* λ.

Если $\lambda\to-0$, то получающееся решение, изображенное на рис. 300 пунктиром, стремится к бесконечности при любом $x>0$. Этот случай представляет меньший интерес.

В более сложных случаях часто наблюдается аналогичное явление. Пусть, например, рассматривается решение уравнения второго порядка, удовлетворяющее двум начальным или краевым (п. 16) условиям, причем при некотором значении параметра $\lambda=\lambda_0$ порядок уравнения понижается до первого. Тогда бывает, что если решение $y_0(x)$ при $\lambda=\lambda_0$ остается конечным, то оно, как решение уравнения первого порядка, удовлетворяет только одному условию, а другому не обязательно. При λ, близком к λ_0, для решения $y(x)$ имеется «пограничный слой» (ширина которого пропорциональна $|\lambda-\lambda_0|$), на протяжении которого $y(x)$ переходит от этого другого условия к $y_0(x)$. Подобная ситуация может возникнуть и для систем дифференциальных уравнений.

29. Методы улучшения невязки. Эти методы основаны на том, что неизвестная функция ищется в виде, включающем несколько параметров, т. е. в виде

$$y=\varphi(x;\ \lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m). \tag{185}$$

При этом правая часть обычно выбирается так, чтобы для любых значений этих параметров удовлетворялись поставленные начальные или граничные условия. После подстановки выражения (185) в заданное дифференциальное уравнение невязка h, т. е. разность между левой и правой частями, будет содержать эти параметры:

$$h=h(x;\ \lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m).$$

Если бы решение (185) было точным, то невязка h тождественно равнялась бы нулю. Поэтому для нахождения параметров $\lambda_1, \lambda_2, \ldots, \lambda_m$ на невязку накладывают m условий, которые заведомо выполняются для тождественно нулевой функции. Например, можно приравнять $h=0$ при m значениях x — это *метод коллокации*. Можно минимизировать интеграл $\int\limits_a^b h^2\,dx$ на том интервале $a\leqslant x\leqslant b$, на котором строится решение, — это *метод наименьших квадратов*. Можно приравнять нулю интегралы $\int\limits_a^b h\psi_1(x)\,dx$, $\int\limits_a^b h\psi_2(x)\,dx, \ldots$ $\ldots, \int\limits_a^b h\psi_m(x)\,dx$, где $\psi_1(x), \psi_2(x), \ldots, \psi_m(x)$ — какая-либо выбранная система функций, — это *метод моментов*, так как подобные интегралы называются *моментами*.

Чем больше введено параметров λ_i, тем более «гибкой» является формула (185), т. е. тем точнее можно представить этой формулой искомое решение, но тем сложнее получаются вычисления. Большое искусство заключается в том, чтобы правильно предугадать вид искомого решения с помощью формулы, содержащей небольшое число параметров. О правильности результата можно судить, сравнивая результаты повторных вычислений по разным методам или с разным числом параметров и т. п.

Если правая часть формулы (185) удовлетворяет не всем поставленным начальным или граничным условиям, то требование, чтобы эти условия удовлетворялись, соответственно уменьшает число условий, накладываемых на невязку.

Рассмотрим простой пример, в котором возможно сравнение с точным решением. Пусть надо решить краевую задачу

$$y''+y=0 \quad (0\leqslant x\leqslant 1), \quad y(0)=0, \quad y(1)=1.$$

Будем искать решение в виде

$$y=\lambda x+\mu x^2. \tag{186}$$

При этом первое граничное условие удовлетворяется автоматически, а второе дает $\lambda+\mu=1$, откуда $y=\lambda x+(1-\lambda)x^2$, и у нас остается всего одна степень свободы, т. е. возможность поставить лишь одно условие для улучшения невязки, которая равна

$$h=y''+y=2(1-\lambda)+\lambda x+(1-\lambda)x^2.$$

Коллокация при $x=\frac{1}{2}$ дает значение $\lambda=\frac{9}{7}$; метод наименьших квадратов для интервала $0\leqslant x\leqslant 1$ дает значение $\lambda=\frac{257}{202}$; метод моментов с функцией $\psi(x)\equiv 1$ дает значение $\lambda=\frac{14}{11}$ (проверьте!). Подстановка этих значений в формулу (186) дает приближенные решения, которые неплохо аппроксимируют точное $y=\frac{\sin x}{\sin 1}$: например, при $x=0{,}5$ оно равно 0,5699, тогда как приближенные решения равны соответственно 0,5714, 0,5681 и 0,5682, ошибка $\pm 0{,}3\%$.

30. Метод упрощения. Этот метод широко применяется на практике, особенно при грубых прикидочных расчетах. Он состоит в том, что само исходное уравнение упрощается путем отбрасывания сравнительно малых членов, замены медленно меняющихся коэффициентов постоянными и т. п. После такого упрощения может получиться уравнение одного из интегрируемых типов и, интегрируя, мы получим функцию, которая может считаться приближенным решением исходного, полного уравнения; во всяком случае, она часто правильно передает характер поведения точного решения. Найдя это «нулевое приближение», иногда удается с его помощью внести поправки, учитывающие упрощение, и тем самым найти «первое приближение» и т. д.

Если уравнение содержит параметры (например, массы, линейные размеры исследуемых объектов и т. п.), то нужно иметь в виду, что при одних значениях этих параметров относительно малыми могут быть одни члены уравнения, а при других значениях—другие, так что упрощение будет при разных значениях параметров производиться по-разному. Кроме того, иногда приходится разбивать интервал изменения независимой переменной на части, в каждой из которых упрощение проводится по-своему.

Особенно полезно такое упрощение уравнения в случаях, когда при самом выводе (написании) дифференциального уравнения делались существенные упрощающие предположения или когда точность, с которой известны рассматриваемые величины, невелика. Так, члены уравнения, меньшие допускаемой погрешности в других его членах, надо безусловно отбросить.

Рассмотрим, например, задачу

$$y''+\frac{1}{1+0{,}1x}y+0{,}2y^3=0,\quad y(0)=1,\quad y'(0)=0;\qquad 0\leqslant x\leqslant 2. \tag{187}$$

Так как коэффициент при y меняется медленно, заменим этот коэффициент его средним значением (п. XIV.5):

$$\frac{1}{2-0}\int_0^2\frac{1}{1+0{,}1x}\,dx=\frac{1}{2}\left.\frac{\ln(1+0{,}1x)}{0{,}1}\right|_0^2=\frac{\ln 1{,}2}{0{,}2}=0{,}911.$$

Кроме того, сравнительно малое третье слагаемое отбросим.

Получим уравнение $y''+0{,}911y=0$ с решением при данных начальных условиях:

$$y=\cos 0{,}954x. \qquad (188)$$

Вид этого приближенного решения подтверждает правомерность отбрасывания последнего слагаемого в уравнении, поскольку отношение третьего члена ко второму порядка $0{,}2y^2<0{,}2$, и потому первый член должен «почти взаимно уничтожиться» со вторым. Внесем поправку на последнее слагаемое, для чего подставим в него приближенное решение (188), оставив коэффициент осредненным:

$$y''+0{,}911y=-0{,}2\cos^3 0{,}954x=-0{,}05\cos 2{,}86x-0{,}15\cos 0{,}954x$$

(мы применили формулу (VIII.14)). По методам п. 18 получаем при заданном начальном условии (проверьте!)

$$y=0{,}993\cos 0{,}954x-0{,}079x\sin 0{,}954x+0{,}007\cos 2{,}86x.$$

Разница по сравнению с нулевым приближением (188) невелика, так что вывод о значении отдельных слагаемых в уравнении (187) остается в силе; в то же время третий член уравнения (187) внес свой вклад в решение. (Подумайте, как можно внести поправку на непостоянство коэффициента в уравнении (187).)

Подобные рассуждения зачастую не блещут строгостью и иногда приводят к ошибкам, однако если они проводятся в соответствии со здравым смыслом, то все же, и притом довольно часто, дают решение, которым можно пользоваться на практике.

31. Метод Эйлера. Мы переходим к изложению некоторых методов численного интегрирования дифференциальных уравнений. Эти методы применяются, если ни один из описанных выше методов «приближенного интегрирования», т. е. получения приближенных формул для решения, не является достаточно эффективным, в частности, если решение требуется с большой точностью на большом интервале изменения аргумента. Кроме того, эти методы применяются в работе быстродействующих электронных вычислительных машин.

Часто целесообразно комбинировать методы приближенного и численного интегрирования. Например, для уравнения

$$y''+(1+e^{-x})\,y=0$$

при каком-либо заданном начальном условии можно для малых x применить формулу Тейлора (п. 24), при средних x в зависимости от требуемой точности — один из методов численного интегрирования, а при больших x — просто отбросить член e^{-x}.

Мы изложим четыре наиболее известных метода численного интегрирования уравнений первого порядка; эти методы очень просто переносятся на системы уравнений первого порядка, к которым приводятся и уравнения высших порядков. В курсах приближенных вычислений можно найти ряд других методов; особо отметим книги [3], [18], [26], [34].

Метод Эйлера прост и нагляден, хотя и недостаточно практически эффективен. Однако его надо хорошо понять, так как многие важные и эффективные методы в различных разделах математики являются, по существу, его развитием.

Метод Эйлера состоит в непосредственной замене производной в дифференциальном уравнении разностным отношением (п. V.7). Пусть рассматривается начальная задача

$$y'=f(x, y), \quad y(x_0)=y_0. \tag{189}$$

Будем считать для простоты шаг h по x постоянным и обозначим

$$x_0+h=x_1, \quad x_0+2h=x_2, \quad x_0+3h=x_3, \ \ldots,$$

а приближенные значения $y(x_k)$ обозначим y_k. Чтобы найти эти значения, заменим в уравнении производную разностным отношением

$$\frac{\Delta y_k}{\Delta x}=f(x_k, y_k), \text{ т. е. } \frac{y_{k+1}-y_k}{h}=f(x_k, y_k)$$

и

$$y_{k+1}=y_k+f(x_k, y_k)\,h. \tag{190}$$

По последней формуле можно, начиная от y_0 и полагая последовательно $k=0, 1, 2, \ldots$, найти значения

$$y_1=y_0+f(x_0, y_0)\,h, \quad y_2=y_1+f(x_1, y_1)\,h, \ \ldots$$

Метод Эйлера имеет простой геометрический смысл, показанный на рис. 301, где изображены также интегральные линии. Он состоит в том, что через заданную точку M_0 мы проводим не искомую интегральную линию, которая нам не известна, а отрезок M_0M_1 касательной к этой линии, руководствуясь направлением поля в точке M_0. Через M_1 мы проводим отрезок, руководствуясь направлением поля в M_1, и т. д. Полученная *ломаная Эйлера* приближенно изображает требуемую интегральную линию, которая получилась бы, если бы шаг h был бесконечно малым, т. е. если бы мы непрерывно подправляли направление ломаной.

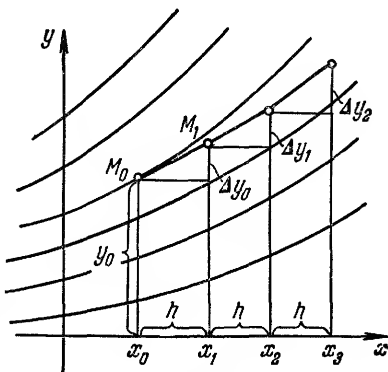


Рис. 301

Легко оценить порядок ошибки в методе Эйлера. Пользуясь формулой (190), мы заменяем приращение решения его дифференциалом $y'_k\Delta x=f(x_k, y_k)\,h$. При этом делается ошибка порядка h^2 (см. формулу (IV.49)). Если мы строим решение на некотором интервале x_0, x и разбиваем его на n частей, то $h=\frac{x-x_0}{n}$, и суммарная ошибка будет иметь порядок $nh^2=\frac{(x-x_0)^2}{n}$. Значит, для повышения точности

в 10 раз, т. е. для вычисления одного дополнительного десятичного знака, требуется увеличить число точек деления также в 10 раз, что значительно увеличит объем вычислительной работы. В этом недостаток метода.

Отметим еще одну особенность метода Эйлера, свойственную и другим методам численного интегрирования дифференциальных уравнений. Мы уже отмечали в п. 7, что решение такого уравнения может, при своем продолжении, обратиться в бесконечность при конечном значении x. В то же время ясно, что решение, построенное по методу Эйлера, остается конечным при всех значениях x. Чтобы правильно передать поведение решения в таких случаях, можно поступить следующим образом: если в результате численного интегрирования будет обнаружено значительное возрастание решения $y(x)$ по абсолютной величине, совершить в дифференциальном уравнении замену вида $y=\frac{1}{z}$. Если тогда при дальнейшем интегрировании окажется, что z переходит через нулевое значение при некотором $x=\alpha$, это и будет означать, что $|y(\alpha)|=\infty$.

32. Метод Ру́нге — Ку́тта. Покажем сначала этот метод, уточняющий метод Эйлера, в более простом варианте. Пусть y_k, т. е. приближенное значение решения при $x=x_k$, уже построено; тогда y_{k+1} можно найти с помощью следующего вычисления:

$$f_k=f(x_k, y_k), \quad \alpha_k=f\left(x_k+\frac{h}{2}, \quad y_k+\frac{f_k h}{2}\right), \quad y_{k+1}=y_k+\alpha_k h. \qquad (191)$$

Геометрический смысл этого вычисления показан на рис. 302. Именно, каждое очередное звено M_kM_{k+1} ломаной, аппроксимирующей интегральную линию, строится так. Сначала проводим отрезок $M_kM'_{k+1}$, руководствуясь направлением f_k поля в точке M_k. Не ограничиваясь этим (как в методе Эйлера), определяем направление α_k поля в середине N_k этого отрезка и проводим отрезок M_kM_{k+1} именно под этим новым направлением. Таким образом, мы производим *пересчет*, уточнение углового коэффициента звеньев ломаной, аппроксимирующей интегральную кривую.

Рис. 302.

Уже из геометрического смысла ясно, что данный метод точнее метода Эйлера, так как здесь учитывается поворот поля на интервале $x_k \leqslant x \leqslant x_{k+1}$.

Это можно подтвердить и вычислением. По формуле Тейлора (п. XII.6)

$$\alpha_k=f(x_k, y_k)+f'_x(x_k, y_k)\frac{h}{2}+f'_y(x_k, y_k)\frac{f_k h}{2}+O(h^2).$$

(под $O(h^2)$ понимается величина, ограниченная по сравнению с h^2, ср.

п. III.11). Подобно п. 24 находим

$$y'_k = f(x_k, y_k) = f_k; \quad y''_k = f'_x(x_k, y_k) + f'_y(x_k, y_k)\, y'_k;$$

$$y_{k+1} = y_k + \alpha_k h = y_k + f(x_k, y_k)\, h + f'_x(x_k, y_k)\frac{h^2}{2} + $$

$$+ f'_y(x_k, y_k)\, f_k \frac{h^2}{2} + O(h^3) = y_k + y'_k h + \frac{y''_k}{2} h^2 + O(h^3). \quad (192)$$

Но точное значение решения при условии $y(x_k) = y_k$ равняется

$$y(x_k + h) = y_k + y'_k h + \frac{y''_k}{2} h^2 + O(h^3) \quad (193)$$

(см. формулу Тейлора (IV.50)). Сравнивая формулы (192) и (193), видим, что значения $y(x_k + h)$ и y_{k+1} могут различаться только в величинах порядка не менее h^3. Отсюда, как в конце п. 31, легко заключить, что суммарная ошибка метода имеет порядок $\frac{1}{n^2}$ или, что то же, порядок h^2. Значит, если число точек деления увеличить в 10 раз, то точность повысится в 100 раз.

Еще более точный результат получится, если вычислять по схеме:

$$f_k = f(x_k, y_k),\ \alpha_k = f\left(x_k + \frac{h}{2},\ y_k + \frac{f_k h}{2}\right),\ \beta_k = f\left(x_k + \frac{h}{2},\ y_k + \frac{\alpha_k h}{2}\right),$$

$$\gamma_k = f(x_k + h,\ y_k + \beta_k h),\ y_{k+1} = y_k + \frac{1}{6}(f_k + 2\alpha_k + 2\beta_k + \gamma_k)\, h.$$

Вычисления, подобные (192), показывают, что ошибка на каждом шаге здесь имеет порядок h^5, а потому суммарная ошибка — порядок h^4. Значит, если число точек деления увеличится в 10 раз, то точность повысится в 10 000 раз.

33. Метод Адамса. Этот метод, предложенный в 1883 г. английским астрономом Дж. Адамсом (1819—1892), основан на второй интерполяционной формуле Ньютона (V.29), которую мы применим для производной $y'(x)$ от решения, начиная от некоторого значения $x_k = x_0 + kh$:

$$y'(x) = y_k + \Delta y'_{k-1}\frac{x - x_k}{h!} + \frac{\Delta^2 y'_{k-2}}{2!}\frac{x - x_k}{h}\left(\frac{x - x_k}{h} + 1\right) + $$

$$+ \frac{\Delta^3 y'_{k-3}}{3!}\frac{x - x_k}{h}\left(\frac{x - x_k}{h} + 1\right)\left(\frac{x - x_k}{h} + 2\right). \quad (194)$$

При этом в указанной формуле мы вместо $t = x_{k+1} - x$ подставили $x_k - x = -(x - x_k)$ и соответственно вместо y_{k+1} — значение y'_k; кроме того, мы заменили знак приближенного равенства знаком точного равенства, хотя, конечно, формула (194) приближенная и ее ошибка имеет порядок $\Delta^4 y'$, т. е. h^4 (п. V.7). Интегрирование формулы (194) от x_k до $x_{k+1} = x_k + h$ дает после подстановки $\frac{x - x_k}{h} = s$

(проверьте!)

$$y_{k+1}=y_k+\left(y'_k+\frac{1}{2}\Delta y'_{k-1}+\frac{5}{12}\Delta^2 y'_{k-2}+\frac{3}{8}\Delta^3 y'_{k-3}\right)h. \qquad (195)$$

Погрешность формулы (195) получается в результате интегрирования погрешности формулы (194), т. е. имеет порядок h^5 (почему?).

Применяется формула (195) следующим образом. Сначала каким-либо способом, например с помощью формулы Тейлора, п. 24, или с помощью метода Рунге—Кутта, п. 32, находим значения

$$y_1=y(x_0+h), \quad y_2=y(x_0+2h) \quad \text{и} \quad y_3=y(x_0+3h).$$

Затем вычисляем соответствующие значения

$$y'_0=f(x_0,\ y_0), \quad y'_1=f(x_1,\ y_1), \quad y'_2=f(x_2,\ y_2), \quad y'_3=f(x_3,\ y_3),$$

с помощью которых находим

$$\Delta y'_0=y'_1-y'_0,\ \Delta y'_1,\ \Delta y'_2,\ \Delta^2 y'_0=\Delta y'_1-\Delta y'_0,\ \Delta^2 y'_1,\ \Delta^3 y'_0=\Delta^2 y'_1-\Delta^2 y'_0.$$

Далее, полагая в формуле (195) $k=3$, вычисляем y_4, а с его помощью $y'_4=f(x_4,\ y_4)$, $\Delta y'_3=y'_4-y'_3$, $\Delta^2 y'_2$, $\Delta^3 y'_1$. Затем, полагая в формуле (195) $k=4$, вычисляем y_5, далее с его помощью $y'_5=f(x_5, y_5)$ и т. д. Вычисления проходят по схеме

x	y	$y'=f(x,y)$	$\Delta y'$	$\Delta^2 y'$	$\Delta^3 y'$
x_{k-3}	y_{k-3}	y'_{k-3}	$\Delta y'_{k-3}$	$\Delta^2 y'_{k-3}$	$\Delta^3 y'_{k-3}$
x_{k-2}	y_{k-2}	y'_{k-2}	$\Delta y'_{k-2}$	$\Delta^2 y'_{k-2}$	$\Delta^3 y'_{k-2}$
x_{k-1}	y_{k-1}	y'_{k-1}	$\Delta y'_{k-1}$	$\Delta^2 y'_{k-1}$	
x_k	y_k	y'_k	$\Delta y'_k$		
x_{k+1}	y_{k+1}	→ y'_{k+1}			

34. Метод Милна. С помощью первой интерполяционной формулы Ньютона (V.27) можно получить еще один метод, который является одним из наиболее эффективных. Мы приведем лишь окончательный результат.

Вычисления в методе Милна (1926) проходят по формулам

$$\left.\begin{aligned}
\bar{y}_{k+1}&=y_{k-3}+\frac{4h}{3}(2y'_{k-2}-y'_{k-1}+2y'_k)\\
&\qquad\qquad (\text{где } y'_i=f(x_i,\ y_i)),\\
\bar{y}'_{k+1}&=f(x_{k+1},\ \bar{y}_{k+1}),\\
y_{k+1}&=y_{k-1}+\frac{h}{3}(y'_{k-1}+4y'_k+\bar{y}'_{k+1})
\end{aligned}\right\}\ (k=3,\ 4,\ 5,\ \ldots). \qquad (196)$$

При этом, как и в методе Адамса, значения y_0, y_1, y_2, y_3 должны быть найдены каким-либо иным способом. После этого, полагая в формулах (196) $k=3$, находим последовательно $\overline{y}_4$, $\overline{y}'_4$, y_4. Затем, полагая $k=4$, находим $\overline{y}_5$, $\overline{y}'_5$, y_5 и т. д. Найденные значения y_4, y_5, y_6, ... и являются приближенными значениями решения $y(x)$ при $x=x_4$, x_5, x_6, ..., где $x_i=x_0+ih$.

Оказывается, что абсолютная погрешность, получающаяся при вычислении y_{k+1} по данному методу, приблизительно равна

$$\frac{|y_{k+1}-\overline{y}_{k+1}|}{29}.$$

Поэтому при вычислениях можно попутно проверять, не выходит ли эта погрешность за рамки принятой степени точности вычислений. Если это где-либо произойдет, то, начиная с соответствующего значения x, надо уменьшить шаг, имея в виду, что суммарная ошибка данного метода имеет порядок h^4.

КРАТНЫЕ ИНТЕГРАЛЫ

§ 1. Определение и основные свойства кратных интегралов

1. Примеры, приводящие к понятию кратного интеграла. Рассмотрим тело (Ω), плотность ρ которого известна, но переменна, т. е. в разных точках различна, и предположим, что нам требуется подсчитать массу m этого тела. Аналогичная задача для линейного распределения массы была решена в пп. XIV.1—2, которые мы советуем сейчас просмотреть.

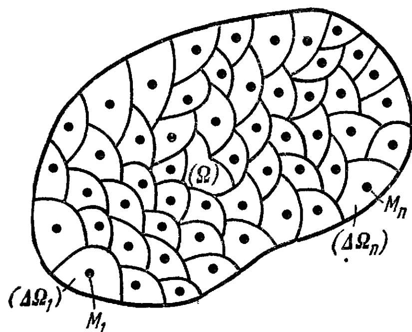


Рис. 303.

Пространственный случай исследуется совершенно аналогично линейному. Для этого разобьем мысленно (Ω) на кусочки—объемчики $(\Delta\Omega_1)$, $(\Delta\Omega_2)$, ..., $(\Delta\Omega_n)$ и выберем в каждом по точке соответственно M_1, M_2, ..., M_n (см. рис. 303, где картину надо представлять себе пространственной; нумерация кусочков производится в произвольном порядке). Если кусочки достаточно малые, то в каждом из них плотность можно без большой ошибки считать постоянной. Тогда массу первого кусочка $m_{(\Delta\Omega_1)}$ можно подсчитать как произведение плотности на объем, т. е. $\rho(M_1)\Delta\Omega_1$, масса второго кусочка находится аналогично и т. д. В целом получаем

$$m_{(\Omega)} \approx \rho(M_1)\Delta\Omega_1 + \rho(M_2)\Delta\Omega_2 + \ldots + \rho(M_n)\Delta\Omega_n = \sum_{k=1}^{n} \rho(M_k)\Delta\Omega_k,$$

где под $\Delta\Omega_k$ понимается объем кусочка $(\Delta\Omega_k)$.

Это равенство приближенное, так как плотность кусочков все же непостоянная; однако оно тем точнее, чем мельче разбиение, и в пределе получаем точное равенство

$$m_{(\Omega)} = \lim \sum_{k=1}^{n} \rho(M_k)\Delta\Omega_k. \qquad (1)$$

Здесь предел берется в процессе безграничного измельчения рассматриваемого разбиения, когда линейные размеры всех кусочков (не просто их объемы!) стремятся к нулю.

Рассуждая аналогичным образом, можно заключить, что если в теле (Ω) распределен заряд с плотностью σ, то сам заряд q можно подсчитать по формуле

$$q = \lim \sum_{k=1}^{n} \sigma(M_k)\,\Delta\Omega_k \tag{2}$$

при таком же смысле обозначений.

Масса или заряд могут быть распределены не по объему, а по поверхности или по линии. Конечно, реально распределение по поверхности означает, что одно из измерений части пространства, занятой массой или зарядом, значительно меньше двух других; аналогично расшифровывается распределение по линии. В этом случае формулы (1) и (2) остаются в силе, если под плотностью понимать поверхностную (т. е. отнесенную к единице площади) или линейную (т. е. отнесенную к единице длины) плотность, а под $\Delta\Omega_k$ понимать соответственно площадь или длину кусочка $(\Delta\Omega_k)$. В общем случае говорят, что $\Delta\Omega_k$ есть *мера области* $(\Delta\Omega_k)$, понимая под этим объем, площадь или длину в зависимости от того, рассматриваются объемные, поверхностные или линейные области.

2. Определение кратных интегралов. Единообразие формул (1) и (2) дает основание для общего определения понятия кратного интеграла. Рассмотрим для определенности объемные интегралы, т. е. под областями будем понимать объемные части пространства, а в качестве меры такой области будем брать ее объем.

Пусть в пространстве задана *конечная* область (Ω) и на ней, т. е. в каждой ее точке M, задана функция $u=f(M)$, принимающая *конечные* значения. Тогда для составления интегральной суммы область (Ω) разбивается на кусочки $(\Delta\Omega_1)$, $(\Delta\Omega_2)$, ..., $(\Delta\Omega_n)$ и в каждой произвольно выбирается точка, соответственно M_1, $M_2, \ldots,$ M_n. Затем составляется сумма

$$\sum_{k=1}^{n} u_k\Delta\Omega_k = \sum_{k=1}^{n} f(M_k)\,\Delta\Omega_k, \tag{3}$$

где под $\Delta\Omega_k$ понимается объем кусочка $(\Delta\Omega_k)$. *Предел этой интегральной суммы в процессе, когда разбиение области* (Ω) *бесконечно измельчается, называется интегралом* (объемным) *от функции* f *по области* (Ω):

$$\int\limits_{(\Omega)} u\,d\Omega = \int\limits_{(\Omega)} f(M)\,d\Omega = \lim \sum_{k=1}^{n} f(M_k)\,\Delta\Omega_k \tag{4}$$

(ср. с основными определениями пп. XIV.2 и XIV.22).

Таким образом, формулы (1) и (2) можно записать так:

$$m=\int\limits_{(\Omega)}\rho d\Omega,\quad q=\int\limits_{(\Omega)}\sigma d\Omega.$$

Как в п. XIV.2, интегрировать можно как непрерывные, так и разрывные функции. При этом существование предела (4) для конечной функции, заданной на конечной области, может быть доказано математически, без ссылки на физический смысл интеграла. (Существенна даже не конечность области, а конечность ее меры, так как нетрудно представить себе область конечной меры, простирающуюся в бесконечность, рис. 264.)

Совершенно аналогично дается определение интеграла по поверхности, плоской или кривой, а также интеграла по линии; конечно, при этом взамен объема кусочка надо взять его площадь или длину. В частности, интеграл по линии — это тот самый криволинейный интеграл по длине дуги, который был разобран в п. XIV.22. Объемные и поверхностные интегралы называются *кратными* по причинам, которые будут ясны из § 3, причем поверхностные интегралы называются *двойными*, а объемные — *тройными*.

3. Основные свойства интегралов. Поскольку основные свойства определенного интеграла пп. XIV.4 и XIV.5 можно вывести из определения п. XIV.2 интеграла как предела интегральной суммы, то эти свойства распространяются и на кратные интегралы. Перечислим эти основные свойства.

1. *Интеграл от суммы равен сумме интегралов*:

$$\int\limits_{(\Omega)}(u_1\pm u_2)\,d\Omega=\int\limits_{(\Omega)}u_1 d\Omega\pm\int\limits_{(\Omega)}u_2 d\Omega.$$

2. *Постоянный множитель можно вынести за знак интеграла*:

$$\int\limits_{(\Omega)}Cu\,d\Omega=C\int\limits_{(\Omega)}u\,d\Omega\qquad(C=\text{const}).$$

3. *Теорема о разбиении области интегрирования*: при любом разбиении области (Ω) на части, скажем, (Ω_1) и (Ω_2), будет

$$\int\limits_{(\Omega)}u\,d\Omega=\int\limits_{(\Omega_1)}u\,d\Omega+\int\limits_{(\Omega_2)}u\,d\Omega.$$

4. *Интеграл от единицы равен мере области интегрирования*:

$$\int\limits_{(\Omega)}d\Omega=\Omega.$$

5. *Если область интегрирования вырождается, так что ее мера становится равной нулю, то и интеграл становится равным нулю.*

В свойствах 4 и 5 мы говорим о мере области (п. 1), понимая под мерой объем, площадь или длину в зависимости от того, рассматриваются объемные, поверхностные или линейные интегралы.

6. Если рассматриваемые переменные размерны, то

$$\left[\int\limits_{(\Omega)} u\,d\Omega\right]=[u]\cdot[\Omega].$$

7. *Случай симметрии*. Если область интегрирования можно разбить на две равные части и если подынтегральная функция в одинаково расположенных точках этих частей принимает равные значения, то интеграл по всей области равен удвоенному интегралу по каждой из этих частей; если же при переходе из одной части в другую подынтегральная функция множится на —1, то интеграл по всей области равен нулю.

Иногда бывает возможно произвести аналогичное разбиение области интегрирования на большее число равных частей, чтобы свести заданный интеграл к интегралу по области более простого вида, чем исходная.

8. *Неравенства можно интегрировать*: если $u_1 \leqslant u_2$, то

$$\int\limits_{(\Omega)} u_1\,d\Omega \leqslant \int\limits_{(\Omega)} u_2\,d\Omega. \tag{5}$$

При этом последнее неравенство может обратиться в равенство, только если $u_1 \equiv u_2$ Впрочем, если это тождество нарушено в точках вырожденной области, имеющей нулевую меру, то интегралы (5) все равно совпадают, так как такое нарушение не влияет на значение интеграла (сравните свойство 5).

9. *Самая грубая оценка интеграла*:

$$u_{\min}\Omega \leqslant \int\limits_{(\Omega)} u\,d\Omega \leqslant u_{\max}\Omega. \tag{6}$$

10. В связи с оценкой (6) находится понятие *среднего* («среднего интегрального», «среднего арифметического») *значения функции u по области* (Ω), которое вводится подобно п. XIV.5, по формуле

$$\int\limits_{(\Omega)} \bar{u}\,d\Omega = \int\limits_{(\Omega)} u\,d\Omega \quad (\bar{u}=\text{const}),$$

т. е.

$$\bar{u}=\frac{1}{\Omega}\int\limits_{(\Omega)} u\,d\Omega, \qquad \int\limits_{(\Omega)} u\,d\Omega=\bar{u}\Omega.$$

Из формул (6) вытекает, что $u_{\min} \leqslant \bar{u} \leqslant u_{\max}$.

Все эти свойства легко иллюстрировать, если под u понимать плотность распределенной массы, а под интегралом — саму массу.

11. *Имеет место неравенство* (см. конец п. XIV.5)

$$\left|\int\limits_{(\Omega)} u\,d\Omega\right| \leqslant \int\limits_{(\Omega)} |u|\,d\Omega.$$

4. Основные методы применения кратных интегралов. Имеются две основные схемы применения кратных интегралов (ср. п. XIV.6). Первая основана на приближенном представлении рассматриваемой величины в виде интегральной суммы (3) с последующим переходом к пределу, как мы это сделали в п. 1. Вторая схема основана на составлении «элемента» (дифференциала) рассматриваемой величины. Остановимся кратко на второй схеме, отложив ее более подробное обоснование до § 2.

Пусть нас интересует значение некоторой величины q, отвечающее (для определенности, объемной) области (Ω), подобно тому как в п. 1 области (Ω) отвечало значение массы или заряда. Тогда мы составляем выражение $dq=\varphi(M)\,d\Omega$, отвечающее бесконечно малому объему $d\Omega$, расположенному в произвольной точке M, которое 1) прямо пропорционально объему $d\Omega$ и 2) отличается от истинного значения Δq на величину высшего порядка малости. Далее, суммируя значения dq по всем «элементам объема» $d\Omega$ в пределах области (Ω), получаем

$$q=q_{(\Omega)}=\int\limits_{(\Omega)}\varphi(M)\,d\Omega. \tag{7}$$

Рассмотрим, например, выражение для *статического момента* материального тела относительно некоторой плоскости (P). Из механики известно, что для конечной системы материальных точек статический момент подсчитывается по формуле

$$S_{(P)}=\sum_k m_k z_k,$$

где m_k — масса k-й точки, а z_k — ее координата, отсчитываемая по оси, перпендикулярной к плоскости (P) (рис. 304).

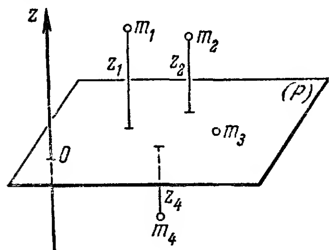


Рис. 304

Если масса распределена по объему (Ω), то, разбив его на частички $(\Delta\Omega_1)$, $(\Delta\Omega_2)$, ..., $(\Delta\Omega_n)$, можно приближенно принять, что масса каждой сосредоточена в одной из ее точек. Отсюда получаем приближенное выражение

$$S_{(P)}\approx\sum_{k=1}^{n}\rho_k z_k\Delta\Omega_k \quad \left(\text{более подробно } \sum_{k=1}^{n}\rho(M_k)\,z(M_k)\,\Delta\Omega_k\right);$$

переходя к пределу, получим

$$S_{(P)}=\int\limits_{(\Omega)}\rho z\,d\Omega. \tag{8}$$

По второму методу надо было бы написать выражение элементарного статического момента

$$dS_{(P)} = \rho z\, d\Omega,$$

откуда, суммируя, мы получили бы ту же формулу (8).

Зная статический момент, легко найти координату *z центра тяжести* рассматриваемого тела:

$$z_{\text{ц.т}} = \frac{S_{(P)}}{m} = \frac{\int\limits_{(\Omega)} \rho z\, d\Omega}{\int\limits_{(\Omega)} \rho\, d\Omega}.$$

Выражение упрощается, если тело однородное, т. е. $\rho = \text{const}$; тогда центр тяжести называется *геометрическим*:

$$z_{\text{ц.т}} = \frac{\rho \int\limits_{(\Omega)} z\, d\Omega}{\rho \int\limits_{(\Omega)} d\Omega} = \frac{1}{\Omega} \int\limits_{(\Omega)} z\, d\Omega. \tag{9}$$

Аналогично находятся другие координаты центра тяжести тела (Ω), а также статические моменты и координаты центра тяжести плоских фигур относительно прямых линий.

5. Геометрический смысл интеграла, взятого по плоской фигуре. Такой интеграл, в отличие от других интегралов, введенных в п. 2, имеет непосредственный геометрический смысл, подобный смыслу простого определенного интеграла (п. XIV.2). Пусть дан интеграл

$$I = \int\limits_{(\Omega)} u\, d\Omega, \tag{10}$$

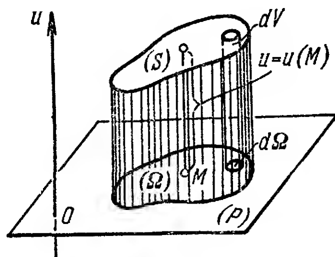


Рис. 305.

где (Ω) — область в некоторой плоскости (P) (рис. 305). Выберем ось u перпендикулярно к плоскости (P) и отложим в каждой точке M области (Ω) значение $u(M)$ подобно тому, как строится график функции двух переменных (рис. 190), причем для простоты будем считать сначала, что $u > 0$. Тогда в пространстве получится поверхность (S) (график подынтегральной функции), которая вместе с областью интегрирования (Ω) и цилиндрической поверхностью, проведенной через контур области (Ω) параллельно оси u, ограничивает некоторое *цилиндрическое тело.*

Геометрический смысл интеграла (10) *состоит в том, что он равен объему указанного цилиндрического тела.* Действительно

элементу $d\Omega$ области (Ω) отвечает объем, который с точностью до малых высшего порядка можно считать цилиндром (рис. 305), т. е. $dV = u\,d\Omega$. Суммируя эти элементарные объемы, т. е. производя интегрирование, получим, что

$$V = \int\limits_{(\Omega)} u\,d\Omega = I,$$

что и утверждалось.

Если u может принимать и отрицательные значения, то, подобно п. XIV.2, объемы, лежащие ниже плоскости (P), войдут в итог со знаком минус.

Аналогично формуле (XIV.28), от объемов цилиндрических тел можно перейти к объему тела произвольной формы

$$V = \int\limits_{(\Omega)} h\,d\Omega,$$

где (Ω) — проекция рассматриваемого тела на некоторую плоскость (P), а $h = h(M)$ — высота сечения этого тела прямой, перпендикулярной к (P) и проходящей через текущую точку M области (Ω).

§ 2. Два типа физических величин

6. Основной пример. Масса и плотность. Рассмотрим материальное тело, вообще говоря, переменной плотности, причем пока не будем принимать во внимание молекулярное строение, а будем считать массу непрерывно распределенной в пространстве. Такое тело имеет в каждой своей точке M определенную плотность ρ, которая является, таким образом, *функцией точки* (п. IX.9), $\rho = \rho(M)$. В отличие от этого *масса не является функцией точки*, так как масса каждой отдельно взятой точки равна нулю Масса является величиной, *распределенной по пространству*; другими словами, каждой мысленно выделенной области (Ω) отвечает значение массы $m_{(\Omega)}$, при этом под (Ω) понимается сама область, т. е. часть пространства, а не просто ее объем.

Связь между массой и плотностью следующая. Пусть для каждой области (Ω) известно соответствующее значение массы $m_{(\Omega)}$. Тогда отношение

$$\frac{m_{(\Omega)}}{\Omega}$$

называется *средней плотностью в* (Ω); здесь под Ω, как и в п. 1, понимается объем области (Ω), т. е. соответствующая размерная величина (можно слово «объем» употреблять и для (Ω) и для Ω, если твердо помнить, о чем идет речь).

Чтобы получить *плотность* в некоторой *точке* M, надо перейти к пределу (ср. п. IV.1), заставив (Ω) стягиваться к точке M:

$$\rho(M) = \lim_{(\Omega)\to M} \rho_{\text{ср}} = \lim_{(\Omega)\to M} \frac{m_{(\Omega)}}{\Omega}; \tag{11}$$

этот процесс аналогичен вычислению производной. Здесь под $(\Omega) \longrightarrow M$ понимается процесс безграничного стягивания (Ω) к точке M; конечно, при этом $\Omega \longrightarrow 0$, но требование $(\Omega) \longrightarrow M$ более сильное, чем требование $\Omega \longrightarrow 0$ (почему?). Итак, плотность в точке — это масса в бесконечно малом объеме, отнесенная к единице этого объема.

Если, обратно, в каждой точке M известно соответствующее значение плотности $\rho(M)$ то массу $m_{(\Omega)}$, отвечающую любой части (Ω) пространства, можно на основании пп. 1—2 найти как интеграл

$$m_{(\Omega)} = \int\limits_{(\Omega)} \rho \, d\Omega.$$

Если учитывать дискретное, молекулярное строение вещества, то в формуле (11) объем (Ω) нельзя даже мысленно «безгранично» стягивать в точку. Взамен формулы (11) надо написать

$$\rho(M) = \frac{m_{(\Delta\Omega)}}{\Delta\Omega},$$

где $(\Delta\Omega)$ — практически бесконечно малый объем (п. III.1), содержащий точку M. Таким образом, плотность реального тела в точке — это средняя плотность по объему, достаточно малому по сравнению с размерами этого тела и в то же время достаточно большому по сравнению с молекулярными размерами. Здесь мы как бы переходим от дискретной картины материального тела к его непрерывной модели, плотность которой получается в результате *осреднения*, т. е. вычисления средней плотности исходной картины по объемам указанных размеров.

Впредь при рассмотрении сплошной среды мы будем рассматривать именно эту непрерывную модель, отвлекаясь от дискретного строения материи.

7. Величины, распределенные по пространству. Имеется ряд физических величин, во многом аналогичных разобранному примеру массы: заряд в диэлектрике, количество тепла, энергия электромагнитного поля и т. п. Их объединяет то, что эти величины *распределены по пространству*. В общем случае мы будем говорить, что *некоторая величина q распределена по пространству, если каждой мысленно выделенной части (Ω) пространства отвечает значение $q_{(\Omega)}$ этой величины*. При этом требуется лишь выполнение *закона сложения* (*закона аддитивности*): при любом разбиении (Ω) на куски, скажем, (Ω_1) и (Ω_2) должно быть $q_{(\Omega)} = q_{(\Omega_1)} + q_{(\Omega_2)}$, т. е. *целое должно равняться сумме частей*.

Величина, распределенная по пространству, имеет в каждой точке определенную плотность: плотность заряда, плотность энергии поля и т. д. В общем случае эта *плотность* φ вводится аналогично формуле (11):

$$\varphi(M) = \lim_{(\Omega) \to M} \frac{q_{(\Omega)}}{\Omega}; \tag{12}$$

отношение, стоящее за знаком предела, является *средней плотностью* величины q в объеме (Ω). Плотность $\varphi = \varphi(M)$ является уже *функцией точки*. Плотность величины q равна значению q, отвечающему бесконечно малой области, «расположенной в точке M», и отнесенному к единице объема.

Обратно, если известна плотность $\varphi(M)$ величины q, то сама величина q находится по методам пп. 1—2:

$$q_{(\Omega)} = \lim \sum_{k=1}^{n} \varphi(M_k) \, \Delta\Omega_k = \int\limits_{(\Omega)} \varphi(M) \, d\Omega, \tag{13}$$

где предел берется в процессе безграничного измельчения разбиения области Ω

В общем случае q и φ могут принимать значения любого знака.

Перепишем формулу (12), написав $(\Delta\Omega)$ вместо Ω, чтобы подчеркнуть малость этого объема:

$$\frac{q_{(\Delta\Omega)}}{\Delta\Omega} \longrightarrow \varphi(M), \quad \frac{q_{(\Delta\Omega)}}{\Delta\Omega} = \varphi(M) + \alpha,$$

где α бесконечно мало́ при $(\Delta\Omega) \longrightarrow M$. Отсюда получаем

$$q_{(\Delta\Omega)} = \varphi(M)\,\Delta\Omega + \alpha\Delta\Omega.$$

Таким образом, значение q, отвечающее малому объему $(\Delta\Omega)$, разбито на две части: одна прямо пропорциональна объему $\Delta\Omega$, а другая имеет высший порядок малости. Поэтому первое слагаемое называется *дифференциалом* или *элементом* величины q (ср. пп. IV. 7—8):

$$dq = \varphi(M)\,\Delta\Omega. \tag{14}$$

Отсюда виден физический смысл dq: это—значение q, которое отвечало бы объему $(\Delta\Omega)$, если бы в нем плотность была постоянной, равной плотности в точке M. На деле Δq, т е. $q_{(\Delta\Omega)}$, не равно dq, а отличается от него на бесконечно малую высшего порядка, т. е. Δq и dq—эквивалентные бесконечно малые (п III.8) Если такими малыми высшего порядка можно пренебречь, то говорят просто, что dq—это значение q, отвечающее бесконечно малому объему $(\Delta\Omega)$, или, просто, бесконечно малая масса, бесконечно малый заряд и т. п.

Если в качестве q рассмотреть саму величину объема, т. е. $q_{(\Omega)} = \Omega$, то соответствующая плотность («объем, отнесенный к единице объема») равна единице, а потому по формуле (14)

$$d\Omega = \Delta\Omega$$

Поэтому формулу (14) можно записать в виде

$$dq = \varphi(M)\,d\Omega, \tag{15}$$

что, как и в п. IV.9, предпочтительнее.

Итак, основные формулы, связывающие величину $q = q_{(\Omega)}$, распределенную в пространстве, и соответствующую функцию точки $\varphi = \varphi(M)$, таковы:

$$\varphi(M) = \left.\frac{dq}{d\Omega}\right|_M, \quad q_{(\Omega)} = \int\limits_{(\Omega)} \varphi(M)\,d\Omega.$$

По этим формулам всегда можно переходить от одной функции к другой. С помощью дифференцирования мы находим плотность, с помощью интегрирования находим распределенную величину по ее плотности.

Отметим, что для того чтобы некоторую величину можно было считать распределенной по пространству, или по поверхности, или по линии, не требуется, чтобы эта величина была «размазана» наподобие массы или заряда. Например, статический момент или момент инерции материального тела удовлетворяют определению, помещенному в начале п. 7, и потому могут считаться величинами, распределенными по объему, хотя они не являются непосредственно «размазанными», а зависят от выбора плоскости или оси отсчета. Поэтому на практике обычно вообще не размышляют на эту тему, а составляют выражение для dq, руководствуясь соображениями, приведенными в п. 4, а затем, опираясь на закон сложения, проводят суммирование (интегрирование) элементов и получают выражение (7).

Величина может быть распределена не по объему, а по поверхности (плоской или кривой) или по линии. В этом случае все результаты этого пункта остаются в силе, если под (Ω) понимать не объемную часть пространства, а часть поверхности (т. е. область на поверхности) или часть линии, а под Ω понимать соответственно площадь или длину этой части, т. е. ее меру (п. 1).

§ 3. Вычисление кратных интегралов в декартовых координатах

8. Интеграл по прямоугольнику. Пусть рассматривается интеграл

$$I = \int\limits_{(\Omega)} u\, d\Omega, \qquad (16)$$

где (Ω) — «координатный» прямоугольник, который в декартовой системе координат описывается неравенствами $a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$ (рис. 306). При составлении интегральной суммы S для интеграла (16) в декартовой системе координат наиболее естественно разбить (Ω) прямыми, параллельными осям координат, после чего применить двойную нумерацию, так что u_{ik} будет значение функции $u = u(x, y)$ в некоторой точке прямоугольника, стоящего в i-м столбце и k-й строке. (Обращаем внимание на то, что эта нумерация не совпадает с той, которая была применена в теории матриц в гл. XI.)

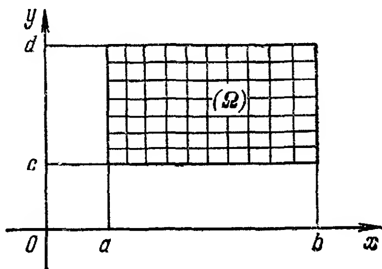


Рис. 306.

Тогда можно написать, что

$$I \approx S = \sum_{i,\,k} u_{ik} \Delta x_i \Delta y_k, \qquad (17)$$

где сумма распространена по всем прямоугольничкам, т. е. по всем значениям i и k, например $i = 1, 2, \ldots, m$, а $k = 1, 2, \ldots, n$.

Сумма вида (17) с двумя индексами суммирования называется *двойной суммой*. Для ее вычисления можно сначала произвести суммирование по k при зафиксированном i, т. е. сложить слагаемые, отвечающие одному (любому) столбцу, а затем результаты просуммировать по i. Тогда получится

$$S = \sum_{i=1}^{m} \left(\sum_{k=1}^{n} u_{ik} \Delta x_i \Delta y_k \right) = \sum_{i=1}^{m} \left(\sum_{k=1}^{n} u_{ik} \Delta y_k \right) \Delta x_i; \qquad (18)$$

здесь мы из *внутренней суммы* вынесли общий множитель Δx_i за скобки. Конечно, такой *переход от двойной суммы к повторной, двукратной*, т. е. от (17) к (18), можно было осуществить и вторым способом: первое, внутреннее, суммирование произвести по i, а второе, внешнее, — по k.

Если рассматриваемое разбиение достаточно мелкое, то последняя скобка в (18) близка к интегралу

$$\sum_{k=1}^{n} u_{ik}\Delta y_k \approx \left(\int_c^d u\,dy\right)_i, \qquad (19)$$

где значок i указывает на то, что значение x берется для i-го столбца. Отсюда

$$S \approx \sum_{i=1}^{m} \left(\int_c^d u\,dy\right)_i \Delta x_i.$$

Но эта сумма — также интегральная сумма, т. е. она близка к интегралу

$$S \approx \int_a^b \left(\int_c^d u(x,\ y)\,dy\right) dx. \qquad (20)$$

С измельчением разбиения равенства (17) и (20) становятся все точнее и точнее и в пределе переходят в точные, так что

$$I = \int_{(\Omega)} u\,d\Omega = \int_a^b \left(\int_c^d u(x,\ y)\,dy\right) dx. \qquad (21)$$

Итак, *для вычисления интеграла по координатному прямоугольнику надо сначала произвести интегрирование по y при фиксированном x в пределах прямоугольника* («внутреннее интегрирование»), *а затем результат, зависящий только от x, проинтегрировать по x в пределах его изменения* («внешнее интегрирование»).

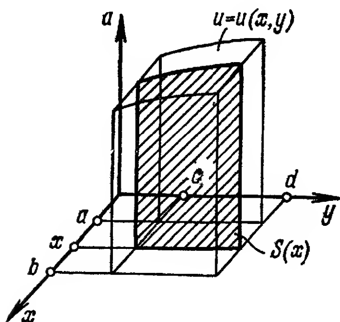


Рис. 307.

При втором способе перехода от двойной суммы (17) к повторной (см. выше) после перехода к пределу мы получили бы

$$\int_{(\Omega)} u\,d\Omega = \int_c^d \left(\int_a^b u(x,\ y)\,dx\right) dy. \qquad (22)$$

Итак, в декартовых координатах, равно как и в любых других (см. § 4), от двойного интеграла можно двумя способами перейти к повторному, двукратному. Об этом следует помнить, так как иногда один способ оказывается более трудным для практического вычисления, а другой — более легким. Переход от одного способа к другому называется *перестановкой порядка интегрирования.*

Формула (21) имеет простой геометрический смысл, показанный на рис. 307. Так как интеграл (16) в силу п. 5 равен объему тела, изображенного на этом рисунке, а объем можно на основе **п. XIV.10**

подсчитывать, интегрируя площадь поперечного сечения, заштрихованного на рисунке, то мы получаем

$$\int_{(\Omega)} u\,d\Omega = V = \int_a^b S(x)\,dx = \int_a^b \left(\int_c^d u\,dy\right) dx.$$

Аналогичный смысл имеет формула (22). Мы привели более сложный вывод этих формул, так как он автоматически распространяется на интегралы любой кратности.

Из-за формул (21) и (22) иногда исходный интеграл (16) обозначают так:

$$I = \iint_{(\Omega)} u\,d\Omega, \qquad I = \iint_{(\Omega)} u\,dx\,dy,$$

имея в виду, что при бесконечно густом разбиении по способу, указанному на рис. 306, будет $d\Omega = dx\,dy$.

Особенно просто вычислить повторный интеграл вида (21) с постоянными пределами интегрирования, если подынтегральная функция представляет собой произведение множителей, каждый из которых зависит только от одной переменной интегрирования: именно, если $u(x, y) = f_1(x) f_2(y)$, то

$$I = \int_a^b \left(\int_c^d f_1(x) f_2(y)\,dy\right) dx = \int_a^b f_1(x) \left(\int_c^d f_2(y)\,dy\right) dx =$$

$$= \int_a^b f_1(x)\,dx \cdot \int_c^d f_2(y)\,dy.$$

Мы получаем произведение однократных интегралов.

9. Интеграл по произвольной плоской фигуре. Пусть (Ω) в интеграле (16) — произвольная фигура в плоскости x, y, например изображенная на рис. 308. Тогда вывод п. 8 переносится с небольшим изменением: именно, взамен интеграла (19) получится интеграл

$$\int_{y_1}^{y_2} u(x, y)\,dy = \int_{\varphi_1(x)}^{\varphi_2(x)} u(x, y)\,dy$$

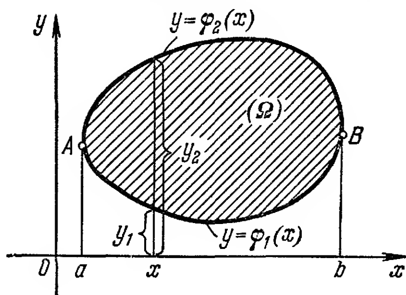


Рис. 308.

(рис. 308); здесь $y = \varphi_1(x)$ и $y = \varphi_2(x)$ — уравнения нижней и верхней частей контура области (Ω), на которые он делится точками A и B. Соответственно и окончательный результат взамен (21) запишется в виде

$$I = \int_{(\Omega)} u\,d\Omega = \int_a^b \left(\int_{\varphi_1(x)}^{\varphi_2(x)} u(x, y)\,dy\right) dx. \qquad (23)$$

Таким образом, пределы интегрирования во внутреннем интеграле в общем случае переменные: они зависят от внешней переменной интегрирования (в данном случае x), причем эта зависимость определяется формой контура. Пределы же у внешнего интеграла постоянные: они определяются общим диапазоном изменения x. Правило, указанное после формулы (21), остается в силе, как видим, и для области (Ω) общего вида.

Можно интегрировать и в другом порядке, сначала по x, а затем по y. Тогда взамен (22) получается формула вида

$$\int\limits_{(\Omega)} u\,d\Omega = \int\limits_{c}^{d}\left(\int\limits_{\psi_1(y)}^{\psi_2(y)} u(x,\ y)\,dx\right)dy \tag{24}$$

(что такое c, d, $\psi_1(y)$, $\psi_2(y)$ на рис. 308?).

Иногда перед расстановкой пределов приходится разбить область интегрирования на части.

Пусть, например, надо переставить порядок интегрирования в интеграле

$$I = \int\limits_0^1 dx \int\limits_{x^2}^{2x} f(x,\ y)\,dy,$$

т. е.

$$I = \int\limits_0^1 \left(\int\limits_{x^2}^{2x} f(x,\ y)\,dy\right)dx. \tag{25}$$

Для этого нужно восстановить область интегрирования. В данном случае она ограничена линиями $x=0$, $x=1$, $y=x^2$ и $y=2x$ (рис. 309), причем первое, внутреннее интегрирование осуществляется по отрезкам, параллельным оси y и показанным на рис. 309 сплошными линиями.

Рис. 309.

После перестановки порядка интегрирования внутреннее интегрирование будет проводиться по отрезкам, параллельным оси x и показанным на рис. 309 пунктиром. Видно, что при $y<1$ интегрирование происходит от прямой до параболы, а при $y>1$ — от прямой до прямой; критическое значение $y=1$ получается из пересечения параболы $y=x^2$ с прямой $x=1$. Поэтому после перестановки порядка интегрирования взамен (25) получим

$$I = \int\limits_0^1 dy \int\limits_{y/2}^{\sqrt{y}} f(x,\ y)\,dx + \int\limits_1^2 dy \int\limits_{y/2}^{1} f(x,\ y)\,dx.$$

В более сложных случаях приходится разбивать область интегрирования на большее число частей. Например, для области, изображенной на рис. 310, при расстановке пределов в декартовых координатах ее пришлось бы разбить на пять частей (каких?).

Приведем простое приложение двукратного интегрирования. Наподобие формулы (9) легко вывести формулы для координат геометрического центра

тяжести плоской фигуры (σ):

$$x_{\text{г.ц.т}} = \frac{\iint\limits_{(\sigma)} x\,dx\,dy}{\sigma}; \quad y_{\text{г.ц.т}} = \frac{\iint\limits_{(\sigma)} y\,dx\,dy}{\sigma}, \tag{26}$$

где под σ понимается площадь фигуры (σ). Пусть фигура (σ) расположена

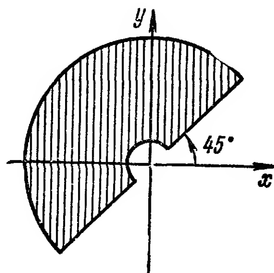


Рис. 310.

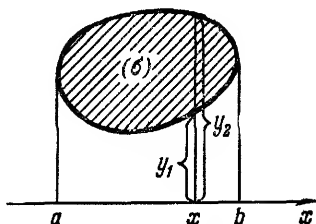


Рис. 311.

по одну сторону от оси x (рис. 311). Тогда вторую формулу (26) можно переписать в виде

$$\sigma \cdot y_{\text{г.ц.т}} = \int\limits_a^b dx \int\limits_{y_1}^{y_2} y\,dy = \int\limits_a^b dx \left(\frac{y_2^2}{2} - \frac{y_1^2}{2}\right) = \frac{1}{2}\int\limits_a^b y_2^2 dx - \frac{1}{2}\int\limits_a^b y_1^2 dx.$$

Умножив обе части на 2π и вспомнив формулу (XIV.35) для объема тела вращения, придем ко *второй теореме Гюльдена: если плоская фигура вращается вокруг оси, лежащей в плоскости этой фигуры и не пересекающей ее, то объем полученного тела вращения равен произведению площади этой фигуры на путь, пройденный ее геометрическим центром тяжести.* На основе этой теоремы легко найти, например, геометрический центр тяжести полукруга (рис. 312):

$$\frac{4}{3}\pi R^3 = \frac{\pi R^2}{2} \cdot 2\pi y_{\text{г. ц. т}},$$

т. е.

$$y_{\text{ц. т}} = \frac{4}{3\pi} R = 0{,}425R.$$

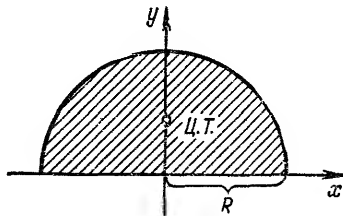


Рис. 312.

10. Интеграл по произвольной поверхности. Рассмотрим интеграл

$$I = \int\limits_{(\Omega)} u\,d\Omega, \tag{27}$$

распространенный по произвольной (вообще говоря, кривой) поверхности (Ω) (рис. 313). Для его вычисления в декартовых координатах надо перейти от фигуры (Ω) к ее проекции на одну из координатных плоскостей, например к проекции (Ω′) на плоскости x, y.

Так как элемент (бесконечно малый участок) кривой поверхности можно с точностью до малых высшего порядка считать плоским, то

$$d\Omega' = d\Omega \cos\alpha = d\Omega \cos(\widehat{\mathbf{n}, z}),$$

где $\mathbf{n}$ — вектор, нормальный к поверхности. Отсюда

$$I = \int\limits_{(\Omega)} u\, d\Omega = \int\limits_{(\Omega')} u \frac{d\Omega'}{\cos(\widehat{\mathbf{n}, z})}. \tag{28}$$

Последний интеграл, распространенный по плоской фигуре (Ω'), вычисляется по методам п. 9.

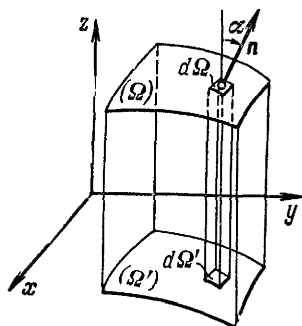


Рис. 313.

Пусть рассматриваемая поверхность задана уравнением $z = f(x, y)$. Тогда из п. XII.2 следует, что вектор

$$\mathbf{n} = -\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k}$$

направлен по нормали к поверхности в любой ее точке x; y; z. Значит (п. VII.10),

$$\cos(\widehat{\mathbf{n}, z}) = \frac{\mathbf{n}\cdot\mathbf{k}}{|\mathbf{n}|\,|\mathbf{k}|} = \frac{1}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}}$$

Поэтому если интеграл (27) задан в форме

$$I = \int\limits_{(\Omega)} u(x, y, z)\, d\Omega,$$

то на основании формулы (28) получаем

$$I = \iint\limits_{(\Omega')} u(x, y, f(x, y)) \sqrt{1 + (f'_x)^2 + (f'_y)^2}\, dx\, dy.$$

В частности, на основе свойства 4 п. 3 получаем формулу для *площади* Ω *произвольной поверхности* (Ω):

$$\Omega = \iint\limits_{(\Omega)} d\Omega = \iint\limits_{(\Omega')} \sqrt{1 + (z'_x)^2 + (z'_y)^2}\, dx\, dy,$$

где, как и выше, (Ω') — это проекция поверхности (Ω) на плоскость x, y.

Иногда при проектировании поверхность (Ω) приходится разбивать на части. Аналогично осуществляется проектирование на плоскости y, z и x, z, если это окажется удобнее. (Выведите формулу для $\cos(\widehat{\mathbf{n}, z})$, если поверхность задана уравнением $F(x, y, z) = 0$.)

11. Интеграл по объему. Рассмотрим теперь интеграл

$$I=\int\limits_{(\Omega)} u\,d\Omega,$$

где (Ω) — некоторое тело, т. е. область в пространстве. Он преобразуется совершенно аналогично тому, как в пп. 8 и 9 преобразовывались интегралы по плоским фигурам. Интегральную сумму, которая оказывается теперь *тройной*, мы представляем в виде *трехкратной*. В самом простом случае, когда (Ω) представляет собой прямоугольный параллелепипед, определенный неравенствами $a\leqslant x\leqslant b$, $c\leqslant y\leqslant d$, $e\leqslant z\leqslant f$, после перехода в интегральной сумме к пределу получим

$$I=\int\limits_a^b dx\int\limits_c^d dy\int\limits_e^f u(x,\ y,\ z)\,dz,$$

т. е.

$$I=\int\limits_a^b\left(\int\limits_c^d\left(\int\limits_e^f u(x,\ y,\ z)\,dz\right)dy\right)dx.$$

Впрочем, возможно произвести интегрирование в пяти других порядках, так как последовательность дифференциалов dx, dy, dz допускает шесть перестановок.

Если область (Ω) интегрирования имеет более общий вид, то и пределы интегрирования будут более сложными. Пусть мы хотим расставить пределы в таком порядке:

$$I=\int\limits_{(\Omega)} u\,d\Omega=$$
$$=\int dx\int dy\int u(x,\ y,\ z)\,dz, \qquad (29)$$

причем область интегрирования имеет вид, изображенный на рис. 314. Поскольку *внутреннее* интегрирование

$$\int u(x,\ y,\ z)\,dz,$$

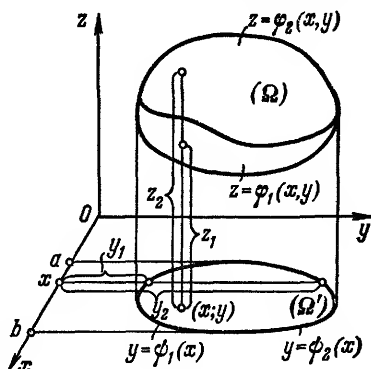


Рис. 314.

которое осуществляется первым, производится по z в пределах области (Ω) при зафиксированных x и y, то пределами для него служат z_1 и z_2 (рис. 314), т. е. $\varphi_1(x, y)$ и $\varphi_2(x, y)$, где $z=\varphi_1(x, y)$ и $z=\varphi_2(x, y)$ — уравнения нижней и верхней частей поверхности тела (Ω).

После интегрирования по z и подстановки пределов результат будет зависеть только от x, y и мы переходим от тела (Ω) к его проекции (Ω') на плоскость x, y. Теперь надо произвести интегрирование по y (это — *среднее* интегрирование) при зафиксированном

x в пределах этой проекции, т. е. от $y_1=\psi_1(x)$ до $y_2=\psi_2(x)$, как это описано в п. 9. Наконец, результат этого второго интегрирования, зависящий уже только от x, надо проинтегрировать в пределах изменения x, т. е. от a до b; это — *внешнее*, последнее интегрирование. Итак, интеграл (29) после расстановки пределов приобретает вид

$$\int\limits_{(\Omega)} u\,d\Omega=\int\limits_a^b dx\int\limits_{\psi_1(x)}^{\psi_2(x)} dy\int\limits_{\varphi_1(x,\,y)}^{\varphi_2(x,\,y)} u(x,\ y,\ z)\,dz.$$

Обратите внимание на то, что у каждого из интегралов пределы могут зависеть только от тех переменных, по которым еще не произведено интегрирование. В частности, пределы внешнего интеграла вообще не могут зависеть от переменных интегрирования.

Подобным образом расставляются пределы и при интегрировании в других порядках. Как и в п. 9, при расстановке пределов для тела (Ω) более сложной формы его иногда приходится разбить на части, чтобы расставить для каждой из них пределы по своему закону.

§ 4. Замена переменных в кратных интегралах

12. Переход к полярным координатам на плоскости. Как и для однократных интегралов, при вычислении двойных интегралов можно вводить различные переменные интегрирования. Мы разберем здесь типичный пример вычисления двойного интеграла в полярных координатах. Пусть рассматривается интеграл

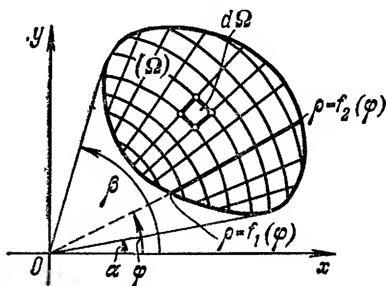


Рис. 315.

$$I=\int\limits_{(\Omega)} u\,d\Omega,$$

где (Ω) — область на плоскости, изображенная на рис. 315. Если требуется произвести интегрирование с помощью полярных координат, то разбиение области надо осуществить посредством координатных линий полярной системы, т. е. линий $\rho=\text{const}$ и $\varphi=\text{const}$ (п. II.5), как на рис. 315. Тогда каждую элементарную площадку с точностью до малых высшего порядка можно считать прямоугольником (почему?) со сторонами $d\rho$ и $\rho d\varphi$, т. е.

$$d\Omega=\rho\,d\rho\,d\varphi.$$

Произведя суммирование по всем площадкам, получим

$$I=\iint\limits_{(\Omega)} u\rho\,d\rho\,d\varphi,$$

причем, конечно, подынтегральная функция u должна быть выражена через ρ и φ. Расставляя пределы подобно п. 9, получим

$$\int\limits_{(\Omega)} u\,d\Omega = \int\limits_{\alpha}^{\beta} d\varphi \int\limits_{f_1(\varphi)}^{f_2(\varphi)} u\rho\,d\rho; \tag{30}$$

смысл пределов интегрирования показан на рис. 315.

Расстановка пределов в полярных координатах особенно удобна для областей, ограниченных координатными линиями полярной системы, так как тогда пределы не только внешнего, но и внутреннего интеграла будут постоянными. Например, для области, изображенной на рис. 310, после расстановки пределов интеграл приобретет вид

$$\int\limits_{\pi/4}^{5\pi/4} d\varphi \int\limits_{r_1}^{r_2} u\rho\,d\rho.$$

13. Переход к цилиндрическим и сферическим координатам в пространстве. Пусть рассматривается интеграл

$$I = \int\limits_{(\Omega)} u\,d\Omega, \tag{31}$$

где (Ω) — область в пространстве. Если требуется произвести интегрирование с помощью цилиндрических координат (п. X.1), то разбиение области надо осуществить посредством координатных поверхностей цилиндрической системы, т. е. поверхностей $\rho = \text{const}$, $\varphi = \text{const}$ и $z = \text{const}$. Тогда каждый элементарный объемчик, изображенный на рис. 316, с точностью до малых высшего порядка можно считать прямоугольным параллелепипедом (почему?) с объемом

$$d\Omega = d\rho \cdot \rho\,d\varphi \cdot dz = \rho\,d\rho\,d\varphi\,dz.$$

Поэтому интеграл (31) примет вид

$$I = \iiint\limits_{(\Omega)} u\rho\,d\rho\,d\varphi\,dz,$$

где еще надо расставить пределы подобно тому, как в п. 11 мы расставили пределы в декартовых координатах.

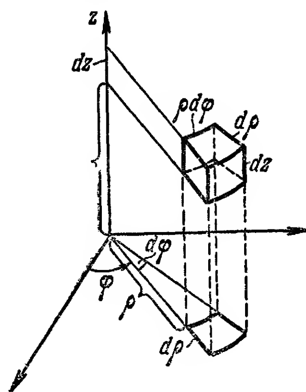


Рис. 316.

Если применяются сферические координаты (п. X.1), то элементарный объемчик, изображенный на рис. 317, и здесь с точностью до малых высшего порядка можно принять за прямоугольный параллелепипед с объемом

$$d\Omega = dr \cdot r\,d\vartheta \cdot r\sin\vartheta\,d\varphi = r^2\sin\vartheta\,dr\,d\vartheta\,d\varphi.$$

Поэтому интеграл (31) примет вид

$$I = \iiint\limits_{(\Omega)} ur^2\sin\vartheta\,dr\,d\vartheta\,d\varphi. \tag{32}$$

Для этих систем координат, как и для любых других, наиболее просто расставлять пределы, если область (Ω) ограничена координатными поверхностями, так как в этом случае пределы внутреннего и среднего интегралов будут постоянными.

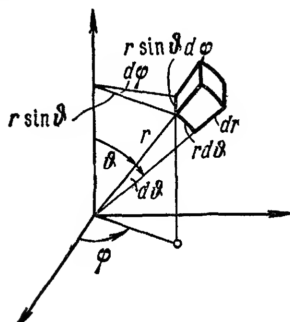


Рис. 317.

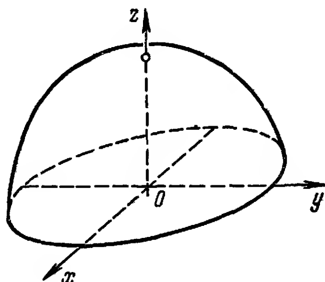


Рис. 318.

Найдем для примера расположение геометрического центра тяжести полушара радиуса R. Для этого расположим полушар, как показано на рис. 318; тогда из симметрии тела ясно, что центр тяжести будет расположен на оси z. Воспользуемся формулой (9) и перейдем к сферическим координатам по формуле (32), заметив, что $z=r\cos\vartheta$:

$$z_{\text{ц.т}}=\frac{1}{{}^2/_3\pi R^3}\times$$

$$\times\iiint\limits_{(\Omega)} r\cos\vartheta\cdot r^2\cdot\sin\vartheta\,dr\,d\vartheta\,d\varphi=$$

$$=\frac{3}{2\pi R^3}\int\limits_0^{2\pi}d\varphi\int\limits_0^{\pi/2}d\vartheta\int\limits_0^{R}r^3\sin\vartheta\times$$

$$\times\cos\vartheta\,dr=\frac{3}{8}R \quad \text{(проверьте!).}$$

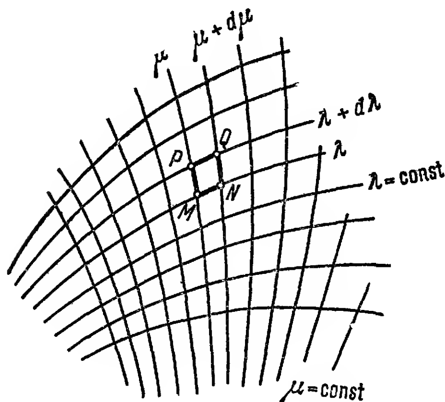


Рис. 319.

14. Общие криволинейные координаты на плоскости. Помимо декартовых и полярных, на плоскости можно ввести много других систем координат. Их общей чертой является то, что *точка на плоскости всегда характеризуется двумя координатами* (п. X.2).

Рассмотрим какую-то общую систему координат λ, μ, для которой координатные линии $\lambda=\text{const}$ и $\mu=\text{const}$ изображены на рис. 319. Если эта сеть нанесена достаточно густо, то плоскость разобьется на фигурки, которые с точностью до малых высшего порядка можно считать параллелограммами.

Пусть линии $\lambda=\text{const}$ проведены через интервал $d\lambda$, а линии $\mu=\text{const}$ — через интервал $d\mu$. Тогда если пренебрегать малыми высшего порядка и обозначить стороны какого-нибудь из «координатных параллелограммчиков» через $ds_\lambda=MP$ и $ds_\mu=MN$ (рис. 319), то эти стороны будут пропорциональны $d\lambda$ и $d\mu$, т. е.

$$ds_\lambda=l_\lambda\,d\lambda,\qquad ds_\mu=l_\mu\,d\mu. \tag{33}$$

Коэффициенты l_λ и l_μ называются *масштабными коэффициентами* или *коэффициентами Ламе* по имени французского математика и инженера Г. Ламé (1795—1870); они дают возможность перейти от координат к линейным размерам. Коэффициенты Ламе для заданной системы координат имеют в различных точках плоскости, вообще говоря, различные значения: например, на рис. 319 они внизу меньше, чем наверху (почему?). Если требуется найти длину конечной дуги координатной линии, то соответствующее равенство (33) нужно проинтегрировать.

Если ввести на плоскости радиус-вектор $\mathbf{r}=\mathbf{r}(\lambda, \mu)=\overrightarrow{OM}$, отсчитываемый от некоторой фиксированной точки O, то стороны $\overrightarrow{MP}$ и $\overrightarrow{MN}$ элементарного параллелограмма на рис. 319 с точностью до малых высшего порядка равны

$$\partial_\lambda \mathbf{r}=\mathbf{r}'_\lambda\, d\lambda \quad \text{и} \quad \partial_\mu \mathbf{r}=\mathbf{r}'_\mu\, d\mu,$$

так как эти приращения радиуса-вектора получаются за счет изменения лишь одной координаты. Отсюда $|\partial_\lambda \mathbf{r}|=|\mathbf{r}'_\lambda|\,d\lambda$; но так как $|d\mathbf{r}|=ds$ (п. VII.23) и потому $|\partial_\lambda \mathbf{r}|=ds_\lambda$, то с помощью (33) получаем

$$l_\lambda=|\mathbf{r}'_\lambda|, \quad l_\mu=|\mathbf{r}'_\mu|.$$

Если на плоскости, помимо криволинейных координат λ, μ, рассматриваются декартовы координаты x, y, то $\mathbf{r}=x\mathbf{i}+y\mathbf{j}$ (п. VII.9) и потому

$$l_\lambda=\left|\frac{\partial x}{\partial \lambda}\mathbf{i}+\frac{\partial y}{\partial \lambda}\mathbf{j}\right|=\sqrt{\left(\frac{\partial x}{\partial \lambda}\right)^2+\left(\frac{\partial y}{\partial \lambda}\right)^2},$$

$$l_\mu=\sqrt{\left(\frac{\partial x}{\partial \mu}\right)^2+\left(\frac{\partial y}{\partial \mu}\right)^2}.$$

(Выведите отсюда с помощью формул $x=\rho\cos\varphi$, $y=\rho\sin\varphi$, а также непосредственно из формул (33), что для полярной системы координат $l_\rho=1$, $l_\varphi=\rho$.)

Площадь $d\sigma$ любого из элементарных параллелограммов, изображенных на рис. 319, пропорциональна как $d\lambda$, так и $d\mu$, т. е.

$$d\sigma=k\, d\lambda\, d\mu, \tag{34}$$

где k — *площадной коэффициент*, также, вообще говоря, различный в разных точках плоскости. Применяя формулу для площади параллелограмма, получим

$$k=\frac{d\sigma}{d\lambda\, d\mu}=\frac{ds_\lambda\, ds_\mu \sin\alpha}{d\lambda\, d\mu}=l_\lambda l_\mu \sin\alpha, \tag{35}$$

где α — угол между координатными линиями.

В частности, *для ортогональных систем координат*, т. е. для систем координат, у которых координатные линии пересекаются под прямым углом,

$$k=l_\lambda\cdot l_\mu. \tag{36}$$

В общем случае из (34) и формулы (VII.21) можно также вывести, что

$$k=\left|\begin{vmatrix}\dfrac{\partial x}{\partial \lambda} & \dfrac{\partial y}{\partial \lambda}\\[2mm] \dfrac{\partial x}{\partial \mu} & \dfrac{\partial y}{\partial \mu}\end{vmatrix}\right|=\left|\frac{D(x, y)}{D(\lambda, \mu)}\right|; \tag{37}$$

по поводу последнего обозначения см. п. IX.13. При этом мы использовали возможность транспонирования определителей (свойство 7 п. VI.2).

Тот же результат можно получить, если заметить, что k в формуле (34) — это коэффициент изменения площадей при переходе от плоскости переменных

λ, μ к плоскости переменных x, y, определяемых формулами $x=x(\lambda, \mu)$, $y=y(\lambda, \mu)$. А этот коэффициент в силу п. XI.14 как раз равен абсолютной величине соответствующего якобиана, т. е. (37).

(Выведите из формул (37), а также (36), что в полярной системе координат $k=\rho$. Чему равны коэффициенты Ламе и площадной коэффициент для декартовой системы координат?)

Если рассматривается интеграл вида

$$I=\int\limits_{(\sigma)} u\, d\sigma,$$

взятый по некоторой плоской области (σ), то на основе формулы (34) получаем

$$I=\iint\limits_{(\sigma)} uk\, d\lambda\, d\mu, \tag{38}$$

где пределы надо расставить наподобие того, как это было сделано для интегралов (23), (24) и (30). Наиболее просто расставлять пределы для области, ограниченной координатными линиями (почему?). (Проверьте, что для полярной системы координат формула (38) переходит в формулу (30).)

15. Общие криволинейные координаты в пространстве. Рассмотрение общих криволинейных координат λ, μ, ν в пространстве проходит совершенно аналогично. Поверхности $\lambda=\text{const}$, $\mu=\text{const}$ и $\nu=\text{const}$ образуют три семейства координатных поверхностей, попарные пересечения которых дают три семейства координатных линий. Координатные поверхности, отвечающие значениям координат λ, $\lambda+d\lambda$; μ, $\mu+d\mu$; ν, $\nu+d\nu$, ограничивают в пространстве объемчик, который с точностью до малых высшего порядка можно принять за параллелепипед, в общем случае косоугольный; для конкретных систем координат он показан на рис. 316 и 317. Одно из ребер этого бесконечно малого параллелепипеда равно

$$ds_\lambda=|\partial_\lambda \mathbf{r}|=\left|\mathbf{r}'_\lambda\right| d\lambda=l_\lambda\, d\lambda,$$

где

$$l_\lambda=\left|\mathbf{r}'_\lambda\right|=\sqrt{\left(\frac{\partial x}{\partial \lambda}\right)^2+\left(\frac{\partial y}{\partial \lambda}\right)^2+\left(\frac{\partial z}{\partial \lambda}\right)^2}$$

называется *масштабным коэффициентом* или *коэффициентом Ламе*. Аналогично выражаются два других ребра параллелепипеда. Его объем выражается по формуле $d\Omega=k\, d\lambda\, d\mu\, d\nu$, где *k — объемный коэффициент*. Поэтому замена переменных в тройном интеграле происходит по формуле

$$\int\limits_{(\Omega)} u\, d\Omega=\iiint\limits_{(\Omega)} uk\, d\lambda\, d\mu\, d\nu. \tag{39}$$

Для ортогональной системы координат $k=l_\lambda l_\mu l_\nu$, тогда как в общем случае вычисление k основано на геометрическом смысле векторно-скалярного произведения (п. VII.15):

$$k=\frac{d\Omega}{d\lambda\, d\mu\, d\nu}=\frac{|(\partial_\lambda \mathbf{r}\times\partial_\mu \mathbf{r})\cdot\partial_\nu \mathbf{r}|}{d\lambda\, d\mu\, d\nu}=\frac{\left|\left(\mathbf{r}'_\lambda\, d\lambda\times\mathbf{r}'_\mu\, d\mu\right)\cdot\mathbf{r}'_\nu\, d\nu\right|}{d\lambda\, d\mu\, d\nu}=$$

$$=\left|\left(\mathbf{r}'_\lambda\times\mathbf{r}'_\mu\right)\cdot\mathbf{r}'_\nu\right|=\left|\begin{vmatrix} \frac{\partial x}{\partial \lambda} & \frac{\partial y}{\partial \lambda} & \frac{\partial z}{\partial \lambda} \\ \frac{\partial x}{\partial \mu} & \frac{\partial y}{\partial \mu} & \frac{\partial z}{\partial \mu} \\ \frac{\partial x}{\partial \nu} & \frac{\partial y}{\partial \nu} & \frac{\partial z}{\partial \nu} \end{vmatrix}\right|=\left|\frac{D(x, y, z)}{D(\lambda, \mu, \nu)}\right|.$$

(Найдите разными способами все указанные коэффициенты для декартовых, цилиндрических и сферических координат.)

16. Координаты на поверхности. На произвольной поверхности можно ввести систему координат (см. п. X.6, в частности, рис. 211). Будем обозначать эти координаты буквами λ, μ. Тогда подобно п. 14 стороны и площадь бесконечно малого координатного параллелограмма с точностью до малых высшего порядка вычисляются по формулам:

$$ds_\lambda = |\partial_\lambda \mathbf{r}| = l_\lambda \, d\lambda,$$

где

$$l_\lambda = |\mathbf{r}'_\lambda| = \sqrt{\left(\frac{\partial x}{\partial \lambda}\right)^2 + \left(\frac{\partial y}{\partial \lambda}\right)^2 + \left(\frac{\partial z}{\partial \lambda}\right)^2}; \qquad ds_\mu = \text{(аналогично)};$$

$d\sigma = k\,d\lambda\,d\mu$, где $k = l_\lambda l_\mu$ для ортогональной системы координат и

$$k = |\mathbf{r}'_\lambda \times \mathbf{r}'_\mu| = \left| \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x}{\partial \lambda} & \frac{\partial y}{\partial \lambda} & \frac{\partial z}{\partial \lambda} \\ \frac{\partial x}{\partial \mu} & \frac{\partial y}{\partial \mu} & \frac{\partial z}{\partial \mu} \end{vmatrix} \right| =$$

$$= \sqrt{\left(\frac{\partial y}{\partial \lambda}\frac{\partial z}{\partial \mu} - \frac{\partial z}{\partial \lambda}\frac{\partial y}{\partial \mu}\right)^2 + \left(\frac{\partial x}{\partial \lambda}\frac{\partial z}{\partial \mu} - \frac{\partial z}{\partial \lambda}\frac{\partial x}{\partial \mu}\right)^2 + \left(\frac{\partial x}{\partial \lambda}\frac{\partial y}{\partial \mu} - \frac{\partial y}{\partial \lambda}\frac{\partial x}{\partial \mu}\right)^2}$$

в общем случае. Переход в интеграле по поверхности к переменным λ, μ осуществляется по формуле (38)

Например, на сфере (т. е поверхности шара) фиксированного радиуса R можно ввести сферические координаты φ, ϑ; получаются как бы обычные сферические координаты в пространстве с зафиксированным значением $r = R$. Эта система координат ортогональна, и из рис. 317 легко вывести, что

$$ds_\varphi = R \sin\vartheta \, d\varphi, \quad ds_\vartheta = R\,d\vartheta,$$

т. е.

$$l_\varphi = R\sin\vartheta, \qquad l_\vartheta = R,$$

и площадной коэффициент k равен $k = l_\varphi l_\vartheta = R^2 \sin\vartheta$. Поэтому интеграл по фигуре (σ), расположенной на сфере, можно вычислять по формуле

$$\int\limits_{(\sigma)} u\,d\sigma = R^2 \iint\limits_{(\sigma)} u \sin\vartheta \, d\varphi \, d\vartheta. \tag{40}$$

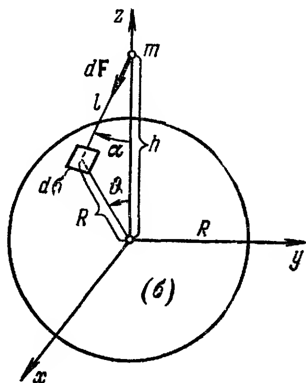


Рис. 320.

Подсчитаем, например, силу притяжения материальной точки массы m материальной сферой (σ) постоянной поверхностной плотности ρ. В силу симметрии задачи можно ограничиться расположением, показанным на рис 320. Каждая элементарная площадка $d\sigma$ притягивает массу m с силой $d\mathbf{F}$, которую можно подсчитать по закону Ньютона:

$$|d\mathbf{F}| = \varkappa \frac{m\rho\,d\sigma}{l^2} = \varkappa \frac{m\rho}{R^2 + h^2 - 2Rh\cos\vartheta}\,d\sigma, \tag{41}$$

где $\varkappa$ — гравитационная постоянная.

При суммировании этих элементарных сил надо суммировать не их модули, так как силы направлены в разные стороны, а проекции этих сил

на определенную ось. Из соображений симметрии ясно, что результирующая сила пойдет по оси z, а потому надо просуммировать проекции всех элементарных сил на ось z:

$$F=\int_{(\sigma)}(d\mathbf{F})_z=\int_{(\sigma)}|d\mathbf{F}|\cos\alpha=\int_{(\sigma)}\varkappa\frac{m\rho}{R^2+h^2-2Rh\cos\vartheta}\,d\sigma\,\frac{l^2+h^2-R^2}{2hl};$$

выражение для $\cos\alpha$ мы нашли по теореме косинусов:

$$R^2=l^2+h^2-2lh\cos\alpha.$$

Подставляя в последнем интеграле $l=\sqrt{R^2+h^2-2Rh\cos\alpha}$ и переходя к сферическим координатам по формуле (40), получим

$$F=\varkappa m\rho\int_{(\sigma)}\frac{h-R\cos\vartheta}{(R^2+h^2-2Rh\cos\vartheta)^{3/2}}\,d\sigma=$$

$$=\varkappa m\rho\int_0^\pi d\vartheta\int_0^{2\pi}\frac{h-R\cos\vartheta}{(R^2+h^2-2Rh\cos\vartheta)^{3/2}}\,R^2\sin\vartheta\,d\varphi=$$

$$=R^2\varkappa m\rho\int_0^\pi\frac{h-R\cos\vartheta}{(R^2+h^2-2Rh\cos\vartheta)^{3/2}}\sin\vartheta\,d\vartheta\int_0^{2\pi}d\varphi=$$

$$=-2\pi R^2\varkappa m\rho\int_0^\pi\frac{h-R\cos\vartheta}{(R^2+h^2-2Rh\cos\vartheta)^{3/2}}\,d\cos\vartheta=$$

$$=2\pi R^2\varkappa m\rho\int_{-1}^{1}\frac{h-Rt}{(R^2+h^2-2Rht)^{3/2}}\,dt\quad\left|\begin{matrix}R^2+h^2-2Rht=l^2\\ l>0\\ -2Rh\,dt=2l\,dl\end{matrix}\right|=$$

$$=2\pi R^2\varkappa m\rho\int_{R+h}^{|R-h|}\frac{h+\frac{l^2-R^2-h^2}{2h}}{l^3}\left(-\frac{2l\,dl}{2Rh}\right)=$$

$$=\frac{\pi R\varkappa m\rho}{h^2}\int_{|R-h|}^{R+h}\left(1+\frac{h^2-R^2}{l^2}\right)dl=$$

$$=\frac{\pi R\varkappa m\rho}{h^2}\left[(R+h)-|R-h|+(h^2-R^2)\left(\frac{1}{|R-h|}-\frac{1}{R+h}\right)\right].\qquad(42)$$

Если $h>R$, то $|R-h|=h-R$, и после подстановки этого выражения в (42) и преобразований получаем (проверьте!)

$$F=\varkappa\frac{m4\pi R^2\rho}{h^2}=\varkappa\frac{mM}{h^2}\qquad(h>R),$$

где M — общая масса сферы. Если же $h<R$, то $|R-h|=R-h$, и аналогично получаем

$$F=0\qquad(h<R).$$

Итак, однородная сфера притягивает точки, расположенные вне ее, с такой силой, как будто вся масса сферы сосредоточена в ее центре, и совсем не притягивает точки, расположенные внутри нее. Пусть теперь дан матери-

альный шар (объемное тело), в котором масса распределена *сферически симметрично*, т. е. плотность зависит лишь от расстояния до центра шара. Такой шар можно представить себе как сумму концентрических бесконечно тонких «пузырей» и для каждого из них применить только что доказанный результат. Тогда мы получим, что этот шар притягивает точку, находящуюся вне его, с такой силой, как будто вся масса шара сосредоточена в его центре; точка же, находящаяся внутри шара, притягивается только его частью, расположенной от центра ближе, чем эта точка. В одном научно-фантастическом романе автор предположил, что внутри земного шара находится концентрическая сферическая полость, поверхность которой населена доисторическими животными. Мы видим, что на этой поверхности сила тяготения, проистекающая из-за наличия в полости атмосферы, была бы направлена к центру Земли, а не к ее оболочке, как думал автор.

§ 5. Варианты кратных интегралов

17. Несобственные интегралы. Теория несобственных кратных интегралов строится подобно тому, как в § XIV.4 это делалось для однократных интегралов. Рассмотрим сначала интеграл

$$I=\int\limits_{(\Omega)} u\,d\Omega, \tag{43}$$

в котором функция u конечная, а область интегрирования бесконечная (неограниченная). Он определяется как предел

$$\int\limits_{(\Omega)} u\,d\Omega=\lim_{(\Omega')\to(\Omega)}\int\limits_{(\Omega')} u\,d\Omega', \tag{44}$$

причем в правой части область (Ω') уже конечная и, расширяясь, стремится исчерпать всю область (Ω) (рис. 321). Если предел (44) существует и конечен, независимо от способа расширения области (Ω'), то интеграл (43) называется сходящимся, в противном случае — расходящимся. Если предел (44) равен бесконечности, то и интеграл (43) расходится к бесконечности.

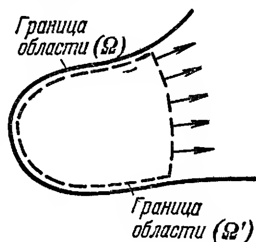


Рис. 321.

Если $u\geqslant 0$, то интеграл (43) либо сходится, либо расходится к $+\infty$. В этом случае для его вычисления можно расставить пределы в любой удобной системе координат по правилам §§ 3—4, причем результат вычисления сам покажет, будет ли интеграл сходящимся (если этот результат конечен) или расходящимся (если он равен бесконечности). Признаки сравнения (XIV.49) и (XIV.50) сохраняют силу. При этом для сравнения применяются как интегралы (XIV.51), так и другие интегралы. Например, если (Ω) — полная плоскость, то часто применяется сравнение с функцией r^{-p}, где $r=\sqrt{x^2+y^2}$ — длина радиуса-вектора. Так как для сходимости существенно поведение подынтегральной функции лишь для больших r, то надо исследовать интеграл

$$\iint\limits_{(r>r_0)} r^{-p}dx\,dy=\int\limits_0^{2\pi}d\varphi\int\limits_{r_0}^{\infty}r^{-p}r\,dr=2\pi\int\limits_{r_0}^{\infty}\frac{1}{r^{p-1}}\,dr \qquad (r_0>0).$$

Согласно п. XIV.15 (формула (XIV.51)) этот интеграл конечен при $p>2$ и бесконечен при $p\leqslant 2$. Аналогично в трехмерном пространстве интеграл от r^{-p} на бесконечности сходится только при $p>3$.

В качестве примера применения несобственных кратных интегралов выведем формулу (XIV 70). Для этого надо исходить из интеграла

$$I=\int_0^\infty\int_0^\infty x^{p-1}y^{p+q-1}e^{-(x+1)y}\,dx\,dy \qquad (p>0,\ q>0),$$

распространенного по первому квадранту плоскости x, y. Так как подынтегральная функция положительна, то интегрирование можно выполнять в любом порядке, что даст

$$1)\ I=\int_0^\infty dy\int_0^\infty x^{p-1}y^{p+q-1}e^{-(x+1)y}\,dx=\left|x=\frac{s}{y}\right|=$$

$$=\int_0^\infty dy\int_0^\infty\left(\frac{s}{y}\right)^{p-1}y^{p+q-1}e^{-\left(\frac{s}{y}+1\right)y}\frac{ds}{y}=$$

$$=\int_0^\infty s^{p-1}e^{-s}\,ds\cdot\int_0^\infty y^{q-1}e^{-y}\,dy=\Gamma(p)\,\Gamma(q);$$

$$2)\ I=\int_0^\infty dx\int_0^\infty x^{p-1}y^{p+q-1}e^{-(x+1)y}\,dy=\left|y=\frac{t}{x+1}\right|=$$

$$=\int_0^\infty dx\int_0^\infty x^{p-1}\left(\frac{t}{x+1}\right)^{p+q-1}e^{-t}\frac{dt}{x+1}=$$

$$=\int_0^\infty\frac{x^{p-1}}{(x+1)^{p+q}}\,dx\cdot\int_0^\infty t^{p+q-1}e^{-t}\,dt=\mathrm{B}(p,q)\,\Gamma(p+q)$$

(см. формулу (XIV.73)). Сравнивая результаты, выводим требуемую формулу.

Если $u\gtrless 0$ и

$$\int_{(\Omega)}|u|\,d\Omega<\infty, \tag{45}$$

то интеграл (43), подобно п. XIV.15, называется *абсолютно сходящимся* и для его вычисления можно расставить пределы в любой удобной системе координат. Если же условие (45) нарушено, то можно проверить, что и интеграл (43) будет расходящимся. Тогда может получиться, что предел (44) будет зависеть от способа расширения области (Ω'), а при расстановке пределов может оказаться, что в одной системе координат результат интегрирования получится конечный, в другой — бесконечный, в третьей — конечный, но отличный от первого, в четвертой получится расходимость колебательного типа (п. XIV.14) и т. д. В этом случае возможность замены переменных и перестановки порядка интегрирования требует специального исследования. Никаких этих проблем не возникает для абсолютно сходящихся интегралов.

Рассмотрение несобственных интегралов иных видов проводится аналогично. Именно, если в области интегрирования имеется точка, или линия и т. д., в которой подынтегральная функция обращается в бесконечность, то эта особенность (точка, линия и т. д.) вырезается из области, после чего вырез произвольным образом стягивается к особенности. Для положительной подынтегральной функции или для функции, положительной всюду вблизи

своих особенностей, можно проводить интегрирование в любой системе координат; для функции произвольного знака это можно делать в случае абсолютной сходимости интеграла. При исследовании *изолированной особенности*, т. е. точки, в которой подынтегральная функция обращается в бесконечность, часто проводится сравнение с интегралами $\iint\limits_{r\leqslant r_0} r^{-p}\,dx\,dy$ на плоскости и $\iiint\limits_{r\leqslant r_0} r^{-p}\,dx\,dy\,dz$ в пространстве. Легко проверить, что первый сходится только при $p<2$, а второй — при $p<3$. Если особенность не изолированная, то условие сходимости можно получить, выбирая систему координат так, чтобы координатные линии шли вдоль особенности.

18. Интегралы, зависящие от параметра. Теория интегралов вида

$$I(\lambda)=\int\limits_{(\Omega)} f(M;\ \lambda)\,d\Omega,$$

где M — точка области (Ω), по которой проводится интегрирование, а λ в процессе интегрирования постоянно, развивается аналогично п. XIV.5; все основные утверждения, доказанные там, остаются справедливыми и здесь. Некоторое затруднение вызывает случай, когда от параметра зависит также область интегрирования. Тогда часто делают предварительную замену переменных в интеграле, после которой область уже остается постоянной; впрочем, можно такие интегралы изучать и непосредственно.

Пусть, например, рассматривается объемный интеграл вида

$$I(\lambda)=\int\limits_{(\varphi_\lambda\leqslant 0)} f(M)\,d\Omega,$$

где заданная функция $\varphi_\lambda(x,\ y,\ z)$ зависит от параметра λ, а интеграл распространен по области, в которой $\varphi_\lambda\leqslant 0$, и требуется подсчитать $dI/d\lambda$. Здесь $dI=I(\lambda+d\lambda)-I(\lambda)$ (с точностью до малых высшего порядка) представляет собой интеграл, взятый по тонкой «пленке», ограниченной поверхностью (S_λ) с уравнением $\varphi_\lambda=0$ и поверхностью $(S_{\lambda+d\lambda})$ с уравнением $\varphi_{\lambda+d\lambda}=0$. Выберем какую-либо точку A на (S_λ), проведем в A нормаль к (S_λ) и будем вдоль этой нормали отсчитывать от A расстояние, считая его положительным в сторону $\varphi_\lambda>0$, т. е. наружу от области интегрирования. Кроме того, обозначим через $\bar{A}$ точку пересечения указанной нормали с $(S_{\lambda+d\lambda})$ и $dn=A\bar{A}$; dn равно ширине «пленки» в точке A. Тогда $\varphi_\lambda(A)=0$ и $\varphi_{\lambda+d\lambda}(\bar{A})=0$ (почему?). Но с точностью до малых высшего порядка

$$\varphi_{\lambda+d\lambda}(\bar{A})=\varphi_\lambda(\bar{A})+\frac{\partial\varphi}{\partial\lambda}\,d\lambda=\varphi_\lambda(A)+|\operatorname{grad}\varphi_\lambda|\cdot dn+\frac{\partial\varphi}{\partial\lambda}\,d\lambda,$$

откуда

$$|\operatorname{grad}\varphi_\lambda|\,dn+\frac{\partial\varphi}{\partial\lambda}\,d\lambda=0,\quad\text{т. е.}\quad dn=-\frac{\partial\varphi/\partial\lambda}{|\operatorname{grad}\varphi_\lambda|}\,d\lambda.$$

Однако для указанной пленки

$$d\Omega=dS\,dn=-\frac{\partial\varphi/\partial\lambda}{|\operatorname{grad}\varphi|}\,dS\,d\lambda;$$

знак dn учитывает, прибавляется эта пленка к исходной области интегрирования или отнимается от нее. Отсюда

$$dI=\int f\cdot\left(-\frac{\partial\varphi/\partial\lambda}{|\operatorname{grad}\varphi|}\,dS\,d\lambda\right),\quad\text{т. е.}\quad\frac{dI}{d\lambda}=-\int\limits_{(\varphi_\lambda=0)} f(M)\,\frac{\partial\varphi/\partial\lambda}{|\operatorname{grad}\varphi|}\,dS.$$

Интеграл может зависеть и от нескольких параметров. В качестве параметра может фигурировать некоторая точка, меняющаяся в пределах какой-либо области. Тогда получаются интегралы вида

$$I(N)=\int\limits_{(\Omega)} f(M;\ N)\,d\Omega_M,$$

где буквы $d\Omega_M$ означают, что при интегрировании точка M является переменной (п. 2), а точка N — постоянной. На такие интегралы распространяются основные свойства п. XIV.5.

При интегрировании интегралов по параметру получаются интегралы высшей кратности. Рассмотрим, например, задачу о вычислении силы $\mathbf{F}$ взаимного притяжения двух материальных тел: (Ω_1) с, вообще говоря, переменной плотностью ρ_1, и (Ω_2) с плотностью ρ_2. Для этого напишем сначала на основе закона Ньютона силу, с которой элемент $d\Omega_1$, расположенный в точке M_1, притягивает элемент $d\Omega_2$, расположенный в точке M_2:

$$d\,d\mathbf{F}=\varkappa\frac{\rho_1\,d\Omega_1\cdot\rho_2\,d\Omega_2}{M_1M_2^2}\,\overrightarrow{M_2M_1^0}=\varkappa\frac{\rho_1\rho_2\overrightarrow{M_2M_1}}{M_1M_2^3}\,d\Omega_1\,d\Omega_2.$$

Интегрируя по (Ω_1), получаем силу, с которой все тело (Ω_1) притягивает элемент $d\Omega_2$:

$$d\mathbf{F}=\varkappa\left(\int\limits_{(\Omega_1)}\frac{\rho_1\overrightarrow{M_2M_1}}{M_1M_2^3}\,d\Omega_1\right)\rho_2\,d\Omega_2.$$

В выписанном интеграле интегрирование производится по M_1 в пределах (Ω_1), фиксированная точка M_2 играет роль параметра. Чтобы получить полную силу притяжения, надо еще проинтегрировать по M_2:

$$\mathbf{F}=\varkappa\int\limits_{(\Omega_2)}\rho_2\,d\Omega_2\int\limits_{(\Omega_1)}\frac{\rho_1\overrightarrow{M_2M_1}}{M_1M_2^3}\,d\Omega_1.$$

Если здесь расставить пределы, то получится *шестикратный* интеграл. Например, в декартовых координатах

$$\mathbf{F}=\varkappa\iiint\limits_{(\Omega_2)}\rho_2(x_2,\ y_2,\ z_2)\,dx_2\,dy_2\,dz_2\times$$
$$\times\iiint\limits_{(\Omega_1)}\frac{\rho_1(x_1,\ y_1,\ z_1)\,[(x_1-x_2)\,\mathbf{i}+(y_1-y_2)\,\mathbf{j}+(z_1-z_2)\,\mathbf{k}]}{[(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2]^{3/2}}\,dx_1\,dy_1\,dz_1;$$

здесь еще надо расставить пределы, руководствуясь формой областей (Ω_1) и (Ω_2).

19. Интеграл по общей мере и обобщенные функции. Рассмотрим для определенности объемные интегралы. В п. 1 мы назвали мерой пространственной области ее объем; однако это — лишь простейший пример меры, называемый иногда *мерой Лебега* по имени французского математика А. Лебéга (1875—1941), который исследовал ее наиболее полно. Возможны и другие меры, по которым также можно проводить интегрирование.

Рассмотрим сначала пример. Пусть в пространстве распределена масса m (п. 6). Подобно п. 18 легко вывести, что сила, с которой эта масса действует на точечную массу m_0, расположенную в точке N, равна

$$\mathbf{F}=\varkappa m_0\int\frac{\overrightarrow{NM}}{NM^3}\,dm, \tag{46}$$

где M — переменная точка интегрирования, а интегрирование распространяется по всей части пространства, занятой массой m.

Если масса m распределена «объемно», т. е. масса каждой поверхности, линии, точки равна нулю, то можно ввести плотность (п. 6) и от интеграла (46) перейти к обычному объемному интегралу

$$\int \frac{\overrightarrow{NM}}{NM^3} \rho \, d\Omega. \tag{47}$$

Однако иногда масса концентрируется на отдельных поверхностях, линиях, точках. Тогда при обычном понимании объемной плотности перейти к интегралу (47) нельзя и интеграл (46) надо рассматривать как интеграл «по мере m».

В общем случае определение меры (*общей меры*) μ в пространстве аналогично основному определению п. 7: каждой мысленно выделенной части пространства (т. е. каждому телу, поверхности, линии, точке) (Ω) должно отвечать значение $\mu_{(\Omega)}$, причем требуется выполнение закона сложения При этом мера поверхности, линии, точки уже не обязательно равна нулю. Обычно предпочитают меры $\mu \geqslant 0$, но иногда приходится от этого отказаться; тогда меру лучше представлять себе не как «массу», а как «электрический заряд». Мера может быть определена не только в пространстве, но также и на поверхности или линии.

Определение *интеграла по мере* (он также называется *интегралом Стилтьеса*) также аналогично обычному (п. 2). Если в области (Ω) задана мера μ, а также функция $u(M)$ (M — произвольная точка (Ω)), то по определению

$$\int\limits_{(\Omega)} u \, d\mu = \lim \sum_{k=1}^{n} u(M_k) \mu_{(\Delta\Omega_k)} \tag{48}$$

при естественном смысле обозначений. Такой интеграл всегда существует, если функция u конечна в (Ω) и мера области (Ω) конечна (если $\mu \gtrless 0$, то надо требовать, чтобы были конечными положительная и отрицательная составляющие меры (Ω), т. е. чтобы $\int\limits_{(\Omega)} |d\mu| < \infty$); впрочем, если функция u разрывная, то приходится уточнить вид применяемых в (48) интегральных сумм, на чем мы не будем останавливаться. Несобственные интегралы по мере определяются, как в п. 17. Свойства интеграла (48) аналогичны свойствам, разобранным в п. 3; в свойствах, связанных с интегрированием неравенств, надо требовать, чтобы $\mu \geqslant 0$.

Если мера каждой поверхности, линии, точки равна нулю, то от интеграла (48) можно перейти к интегралу по объему

$$\int\limits_{(\Omega)} u \, d\mu = \int\limits_{(\Omega)} u \frac{d\mu}{d\Omega} d\Omega = \int\limits_{(\Omega)} u\rho \, d\Omega \qquad \left(\rho = \frac{d\mu}{d\Omega}\right). \tag{49}$$

Такой переход можно совершить и при любой мере, но тогда ρ будет, вообще говоря, *обобщенной функцией*.

Простейшей обобщенной функцией в пространстве является дельта-функция

$$\delta(x-a)\,\delta(y-b)\,\delta(z-c) \tag{50}$$

(п. XIV.25), описывающая плотность единичной массы, расположенной в точке $(a;\ b;\ c)$. Функция $\delta(y-b)\,\delta(z-c)$ описывает плотность линейной массы, расположенной на прямой $y=b$, $z=c$ с единичной линейной плотностью. Функция $\delta(z-c)$ описывает плотность массы, расположенной в плоскости $z=c$ с единичной поверхностной плотностью. С помощью этих и аналогичных

функций, в частности, дельта-функций в криволинейных системах координат, возможно осуществить переход (49) в общем случае.

Свойства обобщенных функций от нескольких переменных аналогичны свойствам обобщенных функций от одной переменной (п. XIV.27). Обобщенную функцию (50) можно применить для построения функции влияния, которая подобно п XIV.26 имеет вид

$$G(M;\ N)=G(x,\ y,\ z;\ \xi,\ \eta,\ \zeta),$$

где (x, y, z) — координаты *точки наблюдения* M, а (ξ, η, ζ) — координаты *источника воздействия* N При рассмотрении процессов, развивающихся во времени, применяется также дельта-функция

$$\delta(x-a)\,\delta(y-b)\,\delta(z-c)\,\delta(t-\tau),$$

которая приводит к функции влияния вида $G(M,\ t;\ N,\ \tau)$.

20. Многомерные интегралы. Меру можно задать и в обобщенном k-мерном пространстве или, как говорят иначе, в *k-мерном многообразии*, определенном в п. X.2. Определение интеграла (48) и его основные свойства полностью сохраняются. Чтобы перейти к повторному интегралу, нужно ввести в пространстве обобщенные координаты (п. X.2) $t_1, \ldots, t_k$, после чего выразить подынтегральную функцию в виде $u=u(t_1, \ldots, t_k)$ и элемент меры $d\mu=\rho(t_1, \ldots, t_k)\,dt_1 dt_2 \ldots dt_k$, отвечающий бесконечно малому *обобщенному координатному параллелепипеду* со сторонами $dt_1, \ldots, dt_k$, расположенному в точке $(t_1, \ldots, t_k)$. Тогда интеграл (48) примет вид

$$\int\limits_{(\Omega)} u\,d\mu=\underbrace{\int\int\ldots\int}_{k\ \text{раз}} u(t_1,\ t_2,\ \ldots,\ t_k)\,\rho(t_1,\ t_2,\ \ldots,\ t_k)\,dt_1\ldots dt_k, \qquad (51)$$

где в правой части надо расставить пределы, руководствуясь изменением координат t в области (Ω).

В формуле (51) ρ, т. е. «плотность меры», понимается как обычная функция, если мера μ каждого многообразия размерности, меньшей k (определенного одним или несколькими уравнениями, связывающими координаты t), равна нулю; так будет, в частности, если плотность меры всюду конечна. В противном случае ρ надо понимать как обобщенную функцию (п. 19).

Если в рассматриваемом пространстве введено понятие объема, то возможно и интегрирование по объему, который является частным случаем меры. Тогда должен быть известен объем бесконечно малого обобщенного координатного параллелепипеда

$$d\Omega=h(t_1,\ \ldots,\ t_k)\,dt_1 dt_2\ldots dt_k, \qquad (52)$$

после чего интеграл $\int\limits_{(\Omega)} u\,d\Omega$ преобразуется подобно (51).

Аналогично вводится понятие интеграла (по мере или по объему) по любому многообразию размерности меньше k в основном k-мерном пространстве. Если в обычном трехмерном пространстве возможны криволинейные, поверхностные и объемные интегралы, то в k-мерном пространстве имеется k типов интегралов (каких?).

В k-мерном числовом пространстве E_k (п. VII.18) в формуле (52) полагают $h\equiv 1$, т. е. за единицу объема принимают объем «единичного k-мерного куба» со сторонами, равными 1. Интегралы низшей кратности в этом пространстве определяют, исходя из того, что p-мерный объем $(1\leqslant p<k)$ p-мерного прямоугольного параллелепипеда, конечного или бесконечно малого, равен произведению его сторон (это — *мера Лебега*).

Подобно п. XIV.23 можно рассматривать *интегралы по координатам*, распространенные по p-мерному $(1\leqslant p\leqslant k)$ многообразию (S) в E_k. При этом

(S) должно быть *ориентированным.* Это понятие при $p>1$ не совсем наглядное и требует уточнения, которое мы сейчас сделаем.

Введем сначала понятие *p-мерного тетраэдра.* По определению одномерным тетраэдром считается отрезок, двумерным — треугольник, трехмерным — треугольная пирамида; чтобы получить четырехмерный тетраэдр, выбирают точку вне трехмерного пространства, в котором расположен трехмерный тетраэдр, и соединяют ее отрезками со всеми точками последнего и т. д. Рассмотрим теперь какой-либо p-мерный тетраэдр с вершинами $A_1, A_2, \ldots$ $\ldots, A_{p+1}$. Ориентация его задается перечислением этих вершин в определенном порядке; при этом считается, что перестановка порядка двух вершин меняет ориентацию на противоположную. Например, для трехмерного тетраэдра с вершинами A, B, C, D порядки $ABCD$ и $DBAC$ определяют одну и ту же ориентацию, а порядок $CBAD$ — противоположную. Каждый тетраэдр можно ориентировать двумя способами.

Если на многообразии (S) произвольно выбрать малый p-мерный тетраэдр, произвольно ориентировать его, а затем перемещать по (S), не меняя его ориентации, то исходная ориентация индуцирует ориентацию всех малых p-мерных тетраэдров на (S), т. е. (S) будет ориентировано. При $p=1$ многообразие (S) является линией и указанный способ равносилен заданию на ней определенного направления; при $p=2$ многообразие (S) является двумерной поверхностью и ориентация равносильна указанию направления обхода контура любой малой фигуры на (S). Если (S) состоит из нескольких кусков, то их ориентацию можно проводить независимо друг от друга.

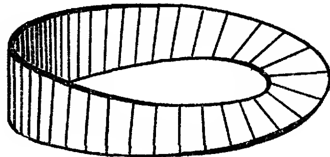


Рис. 322

Следует иметь в виду, что при $p\geqslant 2$ некоторые многообразия ориентировать невозможно. Простейшей неориентируемой поверхностью является *лист Мебиуса* (рис. 322), открытый в 1858 г. немецким геометром А. Мёбиусом (1790—1868).

p-мерный интеграл по координатам, взятый по p-мерному ориентированному многообразию (S) в E_k, определяется так:

$$\int_{(S)}\cdots\int u(t_1, \ldots, t_k)\, dt_{m_1} dt_{m_2} \ldots dt_{m_p} = \lim \sum_{k=1}^{n} u(M_k)\, \Delta S_k', \qquad (53)$$

где в правой части (S) считается разбитым на малые тетраэдры (ΔS_k), ориентированные в соответствии с ориентацией (S), а под $\Delta S_k'$ понимается p-мерный объем проекции (ΔS_k) на плоскость координат $t_{m_1}, t_{m_2}, \ldots, t_{m_p}$, взятый со знаком + или — в зависимости от того, совпадает или нет ориентация этого спроектированного тетраэдра с ориентацией тетраэдра $OC_{m_1} C_{m_2} \ldots C_{m_p}$, где C_m — единичная точка на оси t_m. При этом все индексы $m_1, m_2, \ldots, m_p$ должны быть различными, так как в противном случае интеграл (53) считается равным нулю.

Свойства интеграла (53) аналогичны свойствам интегралов, описанным в п. 3, за исключением свойств, связанных с интегрированием неравенств. При перемене ориентации (S) или при перестановке двух дифференциалов под знаком интеграла он множится на —1 (почему?). Рассматривают также суммы интегралов вида

$$\int_{(S)}\cdots\int \sum_{m_1, \ldots, m_p=1}^{k} u_{m_1, \ldots, m_p}(t_1, \ldots, t_k)\, dt_{m_1} dt_{m_2} \ldots dt_{m_p}. \qquad (54)$$

Частным случаем такого интеграла является интеграл по координатам, взятый по обычной ориентированной двумерной поверхности в обычном трех-

мерном пространстве x, y, z:

$$\iint\limits_{(S)} P(x, y, z)\,dx\,dy + Q(x, y, z)\,dy\,dz + R(x, y, z)\,dx\,dz.$$

При расстановке пределов в интеграле (53) можно выразить в подынтегральной функции через $t_{m_1}, \ldots, t_{m_p}$ значения остальных координат t для точек многообразия (S); затем разбить (S) на части, проекции которых на плоскость $t_{m_1}, \ldots, t_{m_p}$ ориентированы одинаково, после чего расставить пределы по каждой из этих проекций, как в обычном p-мерном интеграле по объему (одни из этих интегралов войдут в итог со знаком $+$, а другие со знаком $-$). Можно также перейти на (S) к каким-либо подходящим криволинейным координатам $s_1, \ldots, s_p$, заменив под интегралом $dt_{m_1} \ldots dt_{m_p}$ на

$$\frac{D\left(t_{m_1}, \ldots, t_{m_p}\right)}{D(s_1, \ldots, s_p)}\,ds_1 \ldots ds_p.$$

§ 6. Векторное поле

Кратные интегралы непосредственно применяются в теории векторного поля. Мы дадим представление об этих применениях в настоящем параграфе, перед чтением которого следует вспомнить определение поля (п. IX.9).

21. Векторные линии. Мы говорим, что задано *векторное поле* **A** (или *поле вектора* **A**), если в каждой точке M пространства или некоторой его области определенс значение векторной величины **A**, т. е. $\mathbf{A}=\mathbf{A}(M)$. Мы будем считать поле стационарным, т. е. не меняющимся с течением времени; если такое изменение имеет место, то наши рассмотрения будут относиться к состоянию поля в любой зафиксированный момент времени. Характерными физическими примерами векторных полей являются поле скоростей **v** или поле *массовых скоростей* $\rho\mathbf{v}$ (ρ — плотность) для потока жидкости или газа, поле сил **F**, поле электрической напряженности **E** и т. д.

Векторной линией поля **A** называется линия (L), в каждой точке которой вектор **A**, отвечающий этой точке, касается (L); другими словами, это — линия, идущая в каждой своей точке вдоль поля. В зависимости от физического смысла поля векторная линия может называться *линией тока* для поля скоростей, *силовой линией* для силового поля и т. д. (Подумайте, почему линии тока совпадают с траекториями частиц жидкости только для стационарных потоков.)

Задача о построении векторных линий заданного векторного поля геометрически равносильна задаче о построении интегральных линий для заданного поля направлений (п. XV.12). Поэтому эта задача сводится к интегрированию системы дифференциальных уравнений; для этого надо ввести в пространство какую-либо систему координат. Если, например, ввести декартовы координаты x, y, z, то вектор **A** можно разложить:

$$\mathbf{A}=\mathbf{A}(x, y, z)=A_x(x, y, z)\,\mathbf{i}+A_y(x, y, z)\,\mathbf{j}+A_z(x, y, z)\,\mathbf{k}. \qquad (55)$$

На основании п. XV.12 систему дифференциальных уравнений векторных линий поля **A** можно записать в симметричной форме

$$\frac{dx}{A_x(x, y, z)}=\frac{dy}{A_y(x, y, z)}=\frac{dz}{A_z(x, y, z)}$$

(ср. уравнения (XV.66)). Для плоских полей (п. IX.9) эта система превращается в уравнение

$$\frac{dx}{A_x(x,\ y)}=\frac{dy}{A_y(x,\ y)}.$$

Как подробно разбиралось в теории дифференциальных уравнений (гл. XV), через каждую неособую точку проходит ровно одна векторная линия. Таким образом, вся область, в которой задано поле, оказывается заполненной векторными линиями. При этом в достаточной близости от любой *обыкновенной*, т. е. неособой точки эти линии напоминают совокупность параллельных отрезков, но несколько искривленных. Вблизи особых точек поля семейство векторных линий может иметь очень сложный вид (см., например, рис. 290).

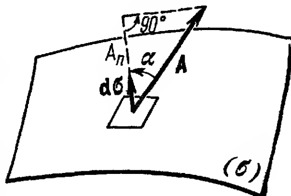


Рис. 323

22. Поток вектора через поверхность. Пусть в пространстве, в котором задано поле вектора **A**, выбрана ориентированная замкнутая или незамкнутая поверхность (σ); напомним, что ориентированность означает указание наружной и внутренней сторон у поверхности (см. п. VII.11). *Потоком вектора* **A** *через поверхность* (σ) называется интеграл по поверхности

$$Q=\int\limits_{(\sigma)} A_n\,d\sigma,$$

где A_n — проекция вектора **A** на внешнюю нормаль к (σ). Применяя вектор площадки (п. VII.11) и скалярное произведение векторов (п. VII.2), можно переписать выражения для потока в виде (рис. 323)

$$Q=\int\limits_{(\sigma)} A\cos\alpha\,d\sigma=\int\limits_{(\sigma)} \mathbf{A}\cdot d\boldsymbol{\sigma}.$$

При вычислении потока, если вектор **A** задан в виде (55), можно воспользоваться преобразованием

$$\mathbf{A}\cdot d\boldsymbol{\sigma}=\mathbf{A}\cdot\mathbf{n}\,d\sigma=(A_x\mathbf{i}+A_y\mathbf{j}+A_z\mathbf{k})\cdot(\cos(\widehat{\mathbf{n},x})\,\mathbf{i}+\cos(\widehat{\mathbf{n},y})\,\mathbf{j}+$$
$$+\cos(\widehat{\mathbf{n},z})\,\mathbf{k})\,d\sigma=A_x\cos(\widehat{\mathbf{n},x})\,d\sigma+A_y\cos(\widehat{\mathbf{n},y})\,d\sigma+A_z\cos(\widehat{\mathbf{n},z})\,d\sigma.$$

Стало быть, интеграл $\int\limits_{(\sigma)}\mathbf{A}\cdot d\boldsymbol{\sigma}$ разбивается на сумму трех интегралов, которые вычисляются аналогично. Например, первый интеграл $\int\limits_{(\sigma)} A_x\cos(\widehat{\mathbf{n},x})\,d\sigma$ можно вычислить, если заметить, что $\cos(\widehat{\mathbf{n},x})\,d\sigma=\pm\,d\sigma_x$, где справа стоит элемент площади проекции (σ_x) поверхности (σ) на плоскость y, z, а знак в правой части определяется знаком $\cos(\mathbf{n},x)$. Если этот знак всюду $+$, то можем написать

$$\int\limits_{(\sigma)} A_x\cos(\widehat{\mathbf{n},x})\,d\sigma=\int\limits_{(\sigma_x)} A_x\,d\sigma_x=\iint\limits_{(\sigma_x)} A_x(x,\ y,\ z)\,dy\,dz,$$

причем в подынтегральную функцию надо подставить выражение $x=x(y,\ z)$ для поверхности (σ). Аналогично рассматривается случай, когда этот знак всюду $-$. Если $\cos(\widehat{\mathbf{n},x})$ на (σ) сменяет знак, то (σ) надо предварительно разбить на куски, на которых этот знак постоянен, после чего преобразовывать интегралы по каждому из этих кусков порознь.

Поток, очевидно, *есть величина скалярная.* Будучи частным случаем интеграла по поверхности, он обладает всеми свойствами таких интегралов (п. 3). Отметим свойство, характерное для потока: при перемене ориентации поверхности поток множится на -1, так как A_n меняет знак. Поток существенно зависит от «направления» поверхности (σ): если поверхность расположена так, что поле пересекает ее всюду изнутри наружу, то $Q>0$, а если всюду снаружи внутрь, то $Q<0$; если же поле частично пересекает (σ) изнутри наружу, а частично снаружи внутрь, то поток равен сумме положительной и отрицательной величин (каких?) и может оказаться положительным, отрицательным или равным нулю. Всегда равен нулю поток через поверхность, целиком заполненную дугами векторных линий, так как для нее поле **A** касается поверхности и потому $A_n=0$.

Физический смысл потока зависит от вида поля. Пусть, например, рассматривается поле **v** скоростей при течении газа. Тогда

$$dQ=\mathbf{v}\cdot\mathbf{d\sigma}$$

равен объему элементарного столба газа, проходящего через площадку ($d\sigma$) за единицу времени (п. VII.11). Поэтому весь поток в этом случае равен объему газа, проносимого через воображаемую поверхность (σ) за единицу времени изнутри наружу. Аналогично легко проверить, что для поля $\mathbf{A}=\rho\mathbf{v}$ массовых скоростей поток равен массе газа, проносимого через (σ) за единицу времени изнутри наружу. (Продумайте на этих примерах свойства потока, указанные в предыдущем абзаце.)

Поток вектора **A** через поверхность (σ) иначе называется *количеством векторных линий* поля **A**, пересекающих (σ) изнутри наружу. Это название несколько условное, так как указанное количество, как правило, имеет размерность и получается дробным, но из-за своей наглядности широко применяется. Надо иметь в виду, что это «количество» понимается в алгебраическом смысле. Так, если одна часть поверхности (σ) пересекается линиями изнутри наружу, а другая часть—снаружи внутрь, то количество линий может быть любого знака или равным нулю в зависимости от того, какую часть пересекает больше линий.

23. Дивергенция. Представим себе в пространстве, где задано поле вектора **A**, некоторый объем (Ω), ограниченный поверхностью (σ), которую мы будем считать ориентированной так, что (Ω) прилегает к ее внутренней стороне. Если поток

$$Q=\oint\limits_{(\sigma)}\mathbf{A}\cdot\mathbf{d\sigma}$$

положителен (кружок у знака интеграла выписывать не обязательно; он подчеркивает, что интеграл берется по замкнутой поверхности), то это значит, что через (σ) наружу проходит больше векторных линий, чем внутрь. Тогда говорят, что в (Ω) имеется *источник* векторных линий, а Q называется *обильностью* этого источника. Если $Q<0$, то говорят, что в (Ω) имеется *сток* или, что то же, источник отрицательной обильности; для простоты мы будем всегда считать сток частным случаем источника. Если $Q=0$, то либо в (Ω) нет ни источников, ни стоков, либо же они взаимно компенсируются. Впрочем, и при $Q\neq0$ в (Ω) могут быть как источники, так и стоки, которые, однако, в этом случае компенсируются не полностью. Представление о линиях, начинающихся в объеме (Ω), оправдывается следующим простым фактом, который мы предоставим доказать читателю: если тело (Ω) с помощью каких-то поверхностей мысленно разбито на несколько тел (Ω_1), (Ω_2), ..., (Ω_k), то поток поля **A** через поверхность тела (Ω) изнутри наружу равен сумме аналогичных потоков, взятых для каждого из тел (Ω_1), (Ω_2), ..., (Ω_k).

Источники векторного поля могут быть сосредоточенными в отдельных точках или быть распределенными по линиям, поверхностям и, наконец, по

пространству (общее понятие о величинах, распределенных по пространству, см. в § 2). Рассмотрим сначала последний случай. Здесь можно говорить не только о средней плотности источника Q/Ω (под Ω понимается численное значение объема (Ω)), но и о плотности источника в любой точке M пространства, равной

$$\lim_{(\Delta\Omega)\to M}\frac{\Delta Q}{\Delta\Omega}=\lim_{(\Delta\Omega)\to M}\left(\int\limits_{(\Delta\sigma)}\mathbf{A}\cdot d\boldsymbol{\sigma}/\Delta\Omega\right), \tag{56}$$

где под $(\Delta\Omega)$ понимается малый объем, содержащий точку M, а под $(\Delta\sigma)$ — его поверхность. Эта плотность источника называется также *дивергенцией* (*расходимостью*) векторного поля $\mathbf{A}$ и обозначается через $\operatorname{div}\mathbf{A}$. Таким образом, можно сказать, что *дивергенция векторного поля — это количество векторных линий, начинающихся в бесконечно малом объеме* (или, что то же, поток поля $\mathbf{A}$ через поверхность этого объема), *в расчете на единицу этого объема.* Отметим, что дивергенция векторного поля есть величина скалярная, более точно, образует скалярное поле, так как она принимает в каждой точке пространства свое значение.

Формулу (56) можно переписать в виде

$$\operatorname{div}\mathbf{A}=\frac{dQ}{d\Omega},\ \text{ т. е. }\ dQ=\operatorname{div}\mathbf{A}\,d\Omega.$$

Получилось выражение для количества векторных линий, начинающихся в элементарном объеме $(d\Omega)$. Производя суммирование (п. 4), получаем выражение для количества векторных линий, начинающихся в конечном объеме (Ω), т. е. для потока вектора $\mathbf{A}$,

$$\oint\limits_{(\sigma)}\mathbf{A}\cdot d\boldsymbol{\sigma}=\int\limits_{(\Omega)}\operatorname{div}\mathbf{A}\,d\Omega, \tag{57}$$

где (Ω) — конечный объем, а (σ) — его поверхность. Эта важная формула называется *формулой Остроградского*, который нашел ее в 1828 г. Она справедлива всегда, если в (Ω) поле $\mathbf{A}$ и его дивергенция не обращаются в бесконечность или если дивергенция обращается в бесконечность, но так, чтобы интеграл в правой части (57) сходился.

Физический смысл дивергенции поля зависит от физического смысла вектора $\mathbf{A}$. Так, если рассматривается поле скоростей $\mathbf{v}$ при течении газа, то $\operatorname{div}\mathbf{v}$ в силу п. 22 равна скорости относительного увеличения бесконечно малого объема, а $\operatorname{div}(\rho\mathbf{v})$ равна плотности источника масс. Например, если в процессе течения газа его масса не меняется (такое изменение может получиться в результате химической или какой-либо подобной реакции), то $\operatorname{div}(\rho\mathbf{v})=0$; в то же время $\operatorname{div}\mathbf{v}>0$, <0 или $=0$ в зависимости от того, будет ли газ в процессе течения расширяться, сжиматься или не менять своей плотности. Для электрического поля $\mathbf{E}$ дивергенция, т. е. $\operatorname{div}\mathbf{E}$, пропорциональна плотности заряда, распределенного в пространстве, и т. д.

Если поле имеет источники, распределенные по линиям или поверхностям (так может быть, только если оно имеет разрывы), то можно говорить о плотности их распределения. В этом случае в правой части формулы (57) должны быть добавлены криволинейные или поверхностные интегралы по линиям и поверхностям, несущим источники и расположенным в (Ω), а также отдельные слагаемые, отвечающие точечным источникам, расположенным в (Ω), если такие имеются. Если пользоваться обобщенными функциями, понятие о которых было дано в п. XIV.7 и в п. 19, то формула (57) справедлива всегда.

В случае плоского поля формула (56) для дивергенции упрощается: именно, в этом случае

$$(\operatorname{div}\mathbf{A})_M = \lim_{(\Delta\sigma)\to M}\left(\oint_{(\Delta l)} A_n\,dl/\Delta\sigma\right), \tag{58}$$

где $(\Delta\sigma)$ — малая фигура в плоскости поля, содержащая точку M, а (Δl) — контур этой фигуры. Как известно (п. IX.9), плоское поле может иметь двоякий физический смысл. Если поле задано только в плоскости, то числитель в правой части (58) *по определению* равен потоку вектора $\mathbf{A}$ через линию (Δl). Если же поле задано в пространстве, но не зависит от одной из декартовых координат, например z, то числитель равен потоку вектора $\mathbf{A}$ через поверхность цилиндра, построенного на $(\Delta\sigma)$, с единичной высотой, параллельной оси z, а знаменатель равен объему этого цилиндра (почему?).

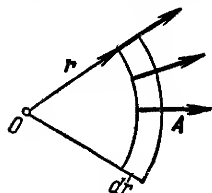


Рис. 324.

Формула Остроградского для плоского поля имеет вид

$$\oint_{(l)} A_n\,dl = \int_{(\sigma)} \operatorname{div}\mathbf{A}\,d\sigma,$$

где (σ) — конечная фигура, а (l) — ее контур.

Иногда оказывается возможным подсчитать дивергенцию непосредственно на основе ее определения (56). Рассмотрим, например, центрально-симметричное поле в пространстве, определенное формулой

$$\mathbf{A} = f(r)\,\mathbf{r}^\circ = \frac{f(r)}{r}\,\mathbf{r},$$

где $\mathbf{r}$ — радиус-вектор текущей точки, а $f(r)$ — некоторая заданная непрерывная функция его модуля (рис. 324). Тогда поток поля через сферу радиуса r равен

$$Q(r) = \int A_n\,d\sigma = \int A_r\,d\sigma = \int f(r)\,d\sigma = f(r)\,4\pi r^2,$$

а потому количество векторных линий, начинающихся в тонком сферическом слое ширины dr, равно

$$dQ = 4\pi d\,[r^2 f(r)] = 4\pi[2rf(r) + r^2 f'(r)]\,dr.$$

Значит, на единицу объема этого слоя приходится

$$\operatorname{div}\mathbf{A} = \frac{dQ}{4\pi r^2\,dr} = \frac{2}{r}f(r) + f'(r).$$

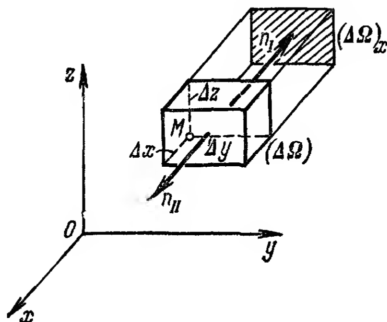


Рис. 325.

24. Выражение дивергенции в декартовых координатах. Пусть в пространстве задана система декартовых координат, так что поле $\mathbf{A}$ можно задать в форме (55). Тогда можно получить удобную формулу для вычисления $\operatorname{div}\mathbf{A}$. Для этого воспользуемся тем, что в формуле (56) для дивергенции форма элементарного объема $(\Delta\Omega)$ несущественна, и выберем в качестве этого объема малый прямоугольный параллелепипед с ребрами, параллельными осям координат (рис. 325). Тогда поток вектора $\mathbf{A}$ через поверхность параллелепипеда, т. е. числитель дроби, стоящей в выражении (56), можно представить в виде шести слагаемых,

отвечающих шести граням параллелепипеда. Рассмотрим сумму двух из этих слагаемых, отвечающих задней и передней граням, которые мы обозначим соответственно цифрами I и II. Тогда (см. рис. 325) $(A_n)_{\rm I}=-(A_x)_{\rm I}$, а на основе формулы Тейлора (п. IV.15)

$$(A_n)_{\rm II}=(A_x)_{\rm II}=(A_x)_{\rm I}+(\partial_x A_x)_{\rm I}+\ldots,$$

где $\partial_x A_x=\frac{\partial A_x}{\partial x}\Delta x$ означает частный дифференциал по x, получающийся из-за того, что точки передней грани отличаются от соответствующих точек задней грани значением координаты x; а многоточием обозначены члены высшего порядка малости. Так как интегрирование по указанным граням сводится к интегрированию по их проекции на плоскость yz, т. е. по $(\Delta\Omega)_x$, то

$$\int\limits_{({\rm I})} A_n\,d\sigma+\int\limits_{({\rm II})} A_n\,d\sigma=\iint\limits_{(\Delta\Omega)_x}(A_n)_{\rm I}\,dy\,dz+\iint\limits_{(\Delta\Omega)_x}(A_n)_{\rm II}\,dy\,dz=$$

$$=\iint\limits_{(\Delta\Omega)_x}\left(\frac{\partial A_x}{\partial x}\right)_{\rm I}\Delta x\,dy\,dz+\ldots=\left(\iint\limits_{(\Delta\Omega)_x}\left(\frac{\partial A_x}{\partial x}\right)_{\rm I}dy\,dz\right)\Delta x+\ldots=$$

$$=\left(\left(\frac{\partial A_x}{\partial x}\right)_{\rm cp}\Delta y\,\Delta z\right)\Delta x+\ldots=\left(\frac{\partial A_x}{\partial x}\right)_M\Delta x\,\Delta y\,\Delta z+\ldots$$

Здесь многоточиями все время обозначаются члены высшего порядка, чем выписанные; в предпоследнем переходе применена формула среднего значения (п. XVI.3, свойство 10), а в последнем переходе—формула $\left(\frac{\partial A_x}{\partial x}\right)_{\rm cp}=\left(\frac{\partial A_x}{\partial x}\right)_M+$бесконечно малая.

Проводя аналогичные вычисления для двух других пар граней, получим выражение для потока через всю поверхность параллелепипеда

$$\int\limits_{(\Delta\sigma)}\mathbf{A}\cdot d\boldsymbol{\sigma}=\left(\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}+\frac{\partial A_z}{\partial z}\right)_M\Delta x\,\Delta y\,\Delta z+\ldots,$$

и так как в данном случае $\Delta\Omega=\Delta x\,\Delta y\,\Delta z$, то

$$\frac{1}{\Delta\Omega}\int\limits_{(\Delta\sigma)}\mathbf{A}\cdot d\boldsymbol{\sigma}=\left(\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}+\frac{\partial A_z}{\partial z}\right)_M+\ldots$$

Переходя к пределу, получаем окончательно

$$\operatorname{div}\mathbf{A}=\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}+\frac{\partial A_z}{\partial z}\;; \qquad (59)$$

мы не пишем индекс M, так как эта формула справедлива в *любой* точке поля.

Для плоского поля в правой части формулы (59) отсутствует третье слагаемое.

25. Линейный интеграл и циркуляция. Пусть в пространстве, в котором задано поле вектора **A**, выбрана ориентированная линия (L), т. е. линия, для которой указано направление ее обхода. Тогда *линейным интегралом вектора* **A** *по линии* (L) называется криволинейный интеграл

$$I=\int\limits_{(L)}A_\tau\,dL, \qquad (60)$$

где A_τ — проекция вектора $\mathbf{A}$ на касательную к (L), проведенную в направлении обхода (рис. 326). Так как вектор $d\mathbf{r}$ идет по $\boldsymbol{\tau}$, а $|d\mathbf{r}|=dL$ (п. VII.23), то выражение для линейного интеграла можно переписать так:

$$I=\int\limits_{(L)} A\cos\alpha\,|d\mathbf{r}|=\int\limits_{(L)} \mathbf{A}\cdot d\mathbf{r}=\int\limits_{(L)} (A_x\,dx+A_y\,dy+A_z\,dz). \qquad (61)$$

Линейный интеграл есть величина скалярная и обладает обычными свойствами криволинейных интегралов (п. XIV.6). При перемене ориентации линии (L) он множится на -1. Если угол α (рис. 326) во всех точках линии (L) острый, то $I>0$, а если он тупой, то $I<0$. Интеграл равен нулю ($I=0$), если угол α все время прямой или, что бывает чаще, если интегралы по частям (L), в которых α острый и α тупой, взаимно уничтожаются.

Рис. 326.

Физический смысл линейного интеграла особенно прост, если $\mathbf{A}$ — поле сил; в этом случае интеграл (60), как мы видели в п. XIV.22, равен работе, совершаемой полем, когда точка, на которую действует сила, проходит линию (L).

Если (L) — замкнутая линия, то линейный интеграл называется *циркуляцией.*

26. Ротор. Нам понадобится выражение для циркуляции по бесконечно малому контуру (ΔL). Для этого допустим, что вектор $\mathbf{A}$ разложен по ортам декартовых осей, т. е. имеет вид (55), а контур (ΔL) расположен вблизи некоторой точки M_0, и вычислим интеграл от первого слагаемого в правой части формулы (61):

$$\oint\limits_{(\Delta L)} A_x(x, y, z)\,dx=\oint\limits_{(\Delta L)} \left[(A_x)_0+\left(\frac{\partial A_x}{\partial x}\right)_0(x-x_0)+\right.$$
$$\left.+\left(\frac{\partial A_x}{\partial y}\right)_0(y-y_0)+\left(\frac{\partial A_x}{\partial z}\right)_0(z-z_0)+\ldots\right]dx. \qquad (62)$$

Здесь мы воспользовались формулой Тейлора (п. XII.6), причем индекс «нуль» говорит о том, что указанные величины берутся в точке M_0, а многоточием обозначены члены высшего порядка малости Однако в п. XIV.23 были доказаны формулы

$$\oint\limits_{(\Delta L)} \left[(A_x)_0+\left(\frac{\partial A_x}{\partial x}\right)_0(x-x_0)-\left(\frac{\partial A_x}{\partial y}\right)_0 y_0-\left(\frac{\partial A_x}{\partial z}\right)_0 z_0\right]dx=$$
$$=\oint\limits_{(\Delta L)} [C_1+C_2x]\,dx=0,$$

$$\oint\limits_{(\Delta L)} y\,dx=-\Delta S\cos(\widehat{\mathbf{n}, z}), \qquad \oint\limits_{(\Delta L)} z\,dx=\Delta S\cos(\widehat{\mathbf{n}, y}),$$

где ΔS — площадь фигуры, ограниченной линией (ΔL) (с точностью до малых высшего порядка эту фигуру можно считать плоской), а $\mathbf{n}$ — нормаль к (ΔS), причем направление этой нормали согласовано с направлением обхода (ΔL) по правилу винта (п. VII.11). Подставляя эти результаты в (62), получим

$$\oint_{(\Delta L)} A_x dx = \left[\left(\frac{\partial A_x}{\partial z}\right)_0 \cos(\widehat{\mathbf{n}, y}) - \left(\frac{\partial A_x}{\partial y}\right)_0 \cos(\widehat{\mathbf{n}, z})\right] \Delta S + \ldots \qquad (63)$$

Вычисляя аналогично два других интеграла в правой части формулы (61), для чего достаточно в формуле (63) провести два раза циклическую перестановку координат, и суммируя результаты, получим

$$\oint_{(\Delta L)} \mathbf{A} \cdot d\mathbf{r} = \left\{\left[\left(\frac{\partial A_x}{\partial z}\right)_0 \cos(\widehat{\mathbf{n}, y}) - \left(\frac{\partial A_x}{\partial y}\right)_0 \cos(\widehat{\mathbf{n}, z})\right] + \right.$$

$$+ \left[\left(\frac{\partial A_y}{\partial x}\right)_0 \cos(\widehat{\mathbf{n}, z}) - \left(\frac{\partial A_y}{\partial z}\right)_0 \cos(\widehat{\mathbf{n}, x})\right] +$$

$$\left. + \left[\left(\frac{\partial A_z}{\partial y}\right)_0 \cos(\widehat{\mathbf{n}, x}) - \left(\frac{\partial A_z}{\partial x}\right)_0 \cos(\widehat{\mathbf{n}, y})\right]\right\} \Delta S + \ldots =$$

$$= \left\{\left[\left(\frac{\partial A_z}{\partial y}\right)_0 - \left(\frac{\partial A_y}{\partial z}\right)_0\right] \cos(\widehat{\mathbf{n}, x}) + \left[\left(\frac{\partial A_x}{\partial z}\right)_0 - \left(\frac{\partial A_z}{\partial x}\right)_0\right] \cos(\widehat{\mathbf{n}, y}) + \right.$$

$$\left. + \left[\left(\frac{\partial A_y}{\partial x}\right)_0 - \left(\frac{\partial A_x}{\partial y}\right)_0\right] \cos(\widehat{\mathbf{n}, z})\right\} \Delta S + \ldots \qquad (64)$$

Полученный результат становится более обозримым, если ввести вектор, называемый *ро́тором* (*вихрем, вихревым вектором*) поля $\mathbf{A}$ и обозначаемый буквами $\operatorname{rot} \mathbf{A}$, по формуле

$$\operatorname{rot} \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) \mathbf{i} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) \mathbf{j} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \mathbf{k}. \qquad (65)$$

Так как этот вектор в каждой точке пространства имеет, вообще говоря, иное значение, то *ротор векторного поля образует новое векторное поле.*

Если заметить дополнительно, что $\mathbf{n}$, как и каждый единичный вектор, имеет разложение (п. VII.9)

$$\mathbf{n} = \cos(\widehat{\mathbf{n}, x}) \mathbf{i} + \cos(\widehat{\mathbf{n}, y}) \mathbf{j} + \cos(\widehat{\mathbf{n}, z}) \mathbf{k},$$

то можно переписать формулу (64) в более простом виде

$$\oint_{(\Delta L)} \mathbf{A} \cdot d\mathbf{r} = (\operatorname{rot} \mathbf{A})_0 \cdot \mathbf{n} \Delta S + \ldots = (\operatorname{rot}_n \mathbf{A})_0 \Delta S + \ldots; \qquad (66)$$

в последнем выражении индекс $\mathbf{n}$ указывает на то, что берется проекция ротора на нормаль $\mathbf{n}$, а многоточием, как и выше, обозначены члены высшего порядка малости. Формула (66) и дает выражение для циркуляции по бесконечно малому контуру.

Если обе части формулы (66) разделить на ΔS, а затем перейти к пределу, то мы получим, при естественных обозначениях,

$$(\operatorname{rot}_n \mathbf{A})_M = \lim_{(\Delta L) \to M} \frac{\oint_{(\Delta L)} \mathbf{A} \cdot d\mathbf{r}}{\Delta S}. \qquad (67)$$

Итак, в любой точке M *проекция ротора поля на любое направление* $\mathbf{n}$ *равна отношению циркуляции поля по бесконечно малому контуру, перпендикулярному к* $\mathbf{n}$, *к площади, охватываемой этим контуром*. Отсюда, в частности, видно, что ротор, определение (65) которого привязано к выбранной системе координат, на самом деле *связан с полем инвариантно*, т. е. не зависит от выбора системы координат, так как правая часть формулы (67) не зависит от такого выбора, а знание проекции вектора на любое направление определяет этот вектор однозначно.

В то же время формула (67) показывает, что ротор является псевдовектором (п.VII.14), так как при перемене правила винта направление обхода (ΔL) надо переменить (п.VII.11), что приведет к перемене знака правой части.

На рис. 327 показано несколько простых примеров векторных полей и указан их ротор, который можно подсчитать по формуле (65) либо же по

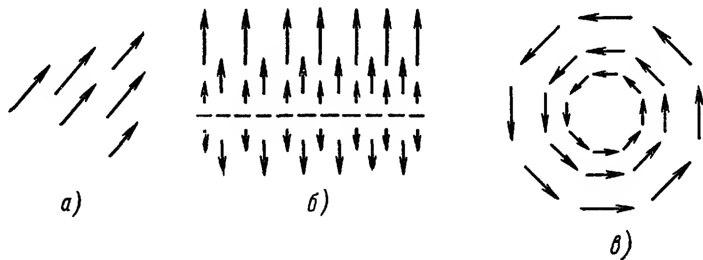


Рис 327.
а) $\mathbf{A}=\text{const}$, $\text{rot}\,\mathbf{A}=0$;
б) $\mathbf{A}=\lambda y\mathbf{j}$, $\text{rot}\,\mathbf{A}=0$,
в) $\mathbf{A}=-\omega y\mathbf{i}+\omega x\mathbf{j}$, $\text{rot}\,\mathbf{A}=2\omega\mathbf{k}$.

формуле (67); для этого (ΔL) надо выбрать соответствующим образом, что мы предоставляем сделать читателю самостоятельно. Третий пример изображает поле скоростей при вращении абсолютно твердого тела вокруг оси z с угловой скоростью ω; из рис. 327, *в* следует, что при таком вращении поле линейных скоростей имеет постоянный ротор, равный удвоенному вектору угловой скорости. Коши показал, что при произвольном движении сплошной среды — газа, жидкости или твердого тела — каждый малый объем участвует одновременно в нескольких движениях, для которых поля скоростей имеют вид, изображенный на рис. 327 (поступательное, деформационное и вращательное движения). Так как ненулевой ротор получается лишь для вращательного движения, то мы видим, что при произвольном движении среды ротор поля линейных скоростей частиц равен в каждой точке удвоенному вектору угловой скорости соответствующей частицы. Конечно, в общем случае ротор получается в различных точках различным. Таким образом, при течении жидкости или газа отличие ротора поля линейных скоростей от нуля указывает на наличие завихренности, чем и объясняется название «ротор».

Особенно простой вид имеет ротор плоского поля $\mathbf{A}=A_x(x, y)\,\mathbf{i}+A_y(x, y)\,\mathbf{j}$: действительно, в силу формулы (65) получаем в этом случае

$$\text{rot}\,\mathbf{A}=\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)\mathbf{k}.$$

27. Формулы Грина и Стокса. Формулы Грина и Стокса осуществляют преобразование циркуляции вектора по замкнутому контуру в двойной интеграл по поверхности, ограниченной этим контуром, причем формула Грина относится к плоскому полю, а формула Стокса — к пространственному. Хотя первая формула непосредственно следует из второй, мы приведем сначала независимый вывод формулы Грина из-за его простоты.

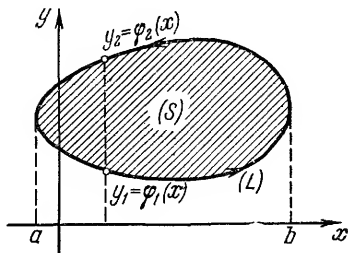


Рис. 328.

Рассмотрим циркуляцию плоского поля $\mathbf{A}=P(x, y)\,\mathbf{i}+Q(x, y)\,\mathbf{j}$ по замкнутому контуру (L), проходимому в положительном направлении и ограничивающему плоскую конечную область (S) (рис. 328). В силу формулы (61) рассматриваемую циркуляцию можно записать в виде

$$\Gamma=\oint_{(L)} P(x, y)\,dx+\oint_{(L)} Q(x, y)\,dy. \tag{68}$$

Первый интеграл равен (см. рис. 328)

$$\int_a^b P(x, y_1)\,dx+\int_b^a P(x, y_2)\,dx=-\int_a^b [P(x, y_2)-P(x, y_1)]\,dx. \tag{69}$$

Под знаком интеграла стоит частное приращение функции P по y, которое можно представить в виде интеграла от производной

$$P(x, y_2)-P(x, y_1)=\int_{\varphi_1(x)}^{\varphi_2(x)} \frac{\partial P}{\partial y}\,dy.$$

Подставляя в (69), получаем

$$-\int_a^b\left(\int_{\varphi_1(x)}^{\varphi_2(x)} \frac{\partial P}{\partial y}\,dy\right)dx=-\iint_{(S)} \frac{\partial P}{\partial y}\,dx\,dy.$$

Аналогичное преобразование второго интеграла (68), которое мы предоставим читателю, и сложение результатов приводят к *формуле Грина*

$$\int_{(L)} (P\,dx+Q\,dy)=\int_{(S)} \left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dS. \tag{70}$$

Этой формулой можно пользоваться, если все функции P, Q, $\frac{\partial Q}{\partial x}$, $\frac{\partial P}{\partial y}$ конечны всюду в (S). Из нее вытекает, в частности, упомянутое в п.XV.6 утверждение о том, что если в односвязной

области (G) удовлетворяется условие $\frac{\partial P}{\partial y}=\frac{\partial Q}{\partial x}$, то в этой области выражение $P\,dx+Q\,dy$ будет полным дифференциалом. В самом деле, тогда в силу формулы (70) для любого замкнутого контура (L) в (G) будет $\oint\limits_{(L)}(P\,dx+Q\,dy)=0$ и наше утверждение вытекает из п. XIV.24. Условие односвязности нужно затем, чтобы для любого такого (L) вся часть плоскости, расположенная внутри (L), также принадлежала (G), что не всегда выполнимо для многосвязной области.

Перейдем теперь к аналогичной формуле в пространстве. Эта формула, найденная в 1854 г. английским физиком и математиком Дж. Стоксом (1819—1903), широко применяется в теории векторного поля. Пусть задан *конечный* ориентированный контур (L), ограничивающий ориентированную поверхность (S), причем ориентации (L) и (S) согласованы (рис. 329). Разобьем (S) на малые площадки (ΔS_1), ..., (ΔS_m) с контурами (ΔL_1), ..., (ΔL_m), которые ориентируем в соответствии с ориентацией (L) и (S). Тогда легко видеть, что

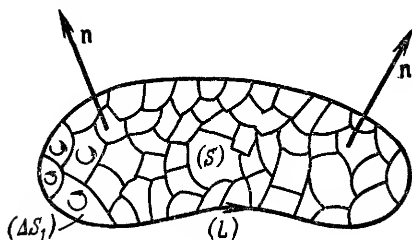


Рис. 329.

$$\oint\limits_{(L)} \mathbf{A}\cdot d\mathbf{r}=\sum_{i=1}^{m}\oint\limits_{(\Delta L_i)} \mathbf{A}\cdot d\mathbf{r}, \qquad (71)$$

так как в правой части интегралы по всем дугам, лежащим внутри (L), взаимно уничтожаются (почему?), а оставшиеся интегралы как раз дают левую часть формулы (71). К отдельным слагаемым в правой части формулы (71) можно применить представление (66), что даст

$$\oint\limits_{(L)} \mathbf{A}\cdot d\mathbf{r}=\sum_{i=1}^{m}(\operatorname{rot}_n \mathbf{A})_i\,\Delta S_i+\ldots,$$

где индекс i показывает, что соответствующие значения относятся к i-й площадке. Сумма в правой части — это интегральная сумма (п. 2), и потому, переходя к пределу при бесконечном измельчении разбиения, получим

$$\oint\limits_{(L)} \mathbf{A}\cdot d\mathbf{r}=\int\limits_{(S)} \operatorname{rot}_n \mathbf{A}\,dS\left(=\int\limits_{(S)} \operatorname{rot}\mathbf{A}\cdot d\mathbf{S}\right). \qquad (72)$$

Итак, *циркуляция поля по замкнутому контуру равна потоку ротора этого поля через поверхность, ограниченную указанным контуром.* Это и есть *формула Стокса.* Она справедлива, если на поверхности (S) поле $\mathbf{A}$ и его ротор конечны, а если ротор и обращается в бесконечность, то так, чтобы интеграл в правой части (72) сходился.

Отметим, что в формуле Стокса контур (L) может состоять из нескольких кусков, тогда они должны быть ориентированы соответственным образом (рис. 330). Это же замечание относится и к формуле Остроградского (п. 23).

Из формулы Стокса вытекает, в частности, упомянутая в п. XIV.24 достаточность условий (XIV.99) для независимости интеграла (XIV.93) от контура интегрирования. Для этого надо рассмотреть поле $\mathbf{A}=P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$, для которого в силу (XIV.99) будет $\operatorname{rot}\mathbf{A}=\mathbf{0}$, а затем, натянув на любой замк-

нутый контур (L) «пленку» (S), получить в силу формулы Стокса равенство (XIV. 95). При этом односвязность области, в которой осуществляются все построения, нужна, чтобы гарантировать возможность такого натягивания пленки: в самом деле, при стягивании контура (L) в точку в пределах области он сам опишет требуемую поверхность (S) (рис. 331)

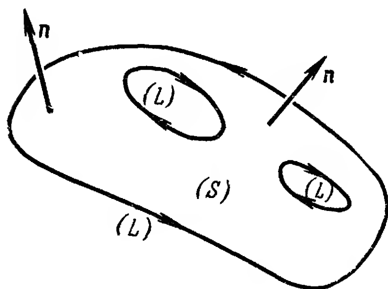


Рис 330.

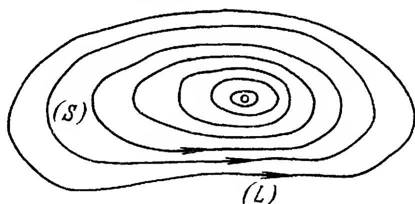


Рис. 331.

28. Выражение векторных операций в криволинейной ортогональной системе координат. Рассмотрим (рис. 332) какую-либо криволинейную *ортогональную* систему координат λ, μ, ν в пространстве. Здесь естественно участвующие векторные поля разлагать в любой точке по системе единичных векторов, касательных к координатным линиям:

$$\mathbf{A} = A_\lambda \mathbf{e}_\lambda + A_\mu \mathbf{e}_\mu + A_\nu \mathbf{e}_\nu.$$

Для выражения градиента скалярного поля u в любой точке M напомним (п. XII.1), что при подсчете градиента с помощью формулы (XII.2) систему декартовых координат можно располагать произвольно, в частности, можно принять $\mathbf{i} = \mathbf{e}_\lambda$, $\mathbf{j} = \mathbf{e}_\mu$, $\mathbf{k} = \mathbf{e}_\nu$. Тогда получим

$$\operatorname{grad} u = \frac{\partial_\lambda u}{\partial s_\lambda} \mathbf{e}_\lambda + \frac{\partial_\mu u}{\partial s_\mu} \mathbf{e}_\mu + \frac{\partial_\nu u}{\partial s_\nu} \mathbf{e}_\nu = \frac{1}{l_\lambda} \frac{\partial u}{\partial \lambda} \mathbf{e}_\lambda + \frac{1}{l_\mu} \frac{\partial u}{\partial \mu} \mathbf{e}_\mu + \frac{1}{l_\nu} \frac{\partial u}{\partial \nu} \mathbf{e}_\nu,$$

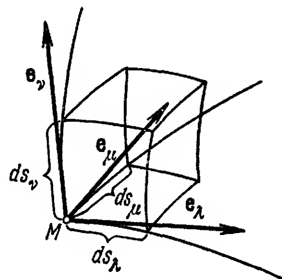


Рис. 332.

где l_λ, l_μ, l_ν — коэффициенты Ламе (п. 15).

При подсчете дивергенции векторного поля нельзя непосредственно воспользоваться формулой (59), так как если принять, как в предыдущем абзаце, $\mathbf{i} = \mathbf{e}_\lambda$ и т. д., то равенство $A_x = A_\lambda$ будет соблюдаться лишь в точке M (почему?) и потому равенство $\partial A_x/\partial x = \partial A_\lambda / l_\lambda \, \partial l_\lambda$, вообще говоря, несправедливо. Здесь можно воспользоваться доказательством, приведенным в начале п. 24, рассмотрев поток поля через поверхность бесконечно малого прямоугольного параллелепипеда, ограниченного координатными линиями (рис. 332). Рассматривая сумму потоков через грани, перпендикулярные к линии λ, получим, что эта сумма с точностью до малых высшего порядка равна

$$\partial_\lambda (A_\lambda \, ds_\mu \, ds_\nu) = \partial_\lambda (l_\mu l_\nu A_\lambda) \, d\mu d\nu = \frac{\partial (l_\mu l_\nu A_\lambda)}{\partial \lambda} \, d\lambda \, d\mu \, d\nu.$$

Суммируя потоки через две другие пары граней и деля на элементарный объем $d\Omega = l_\lambda l_\mu l_\nu\, d\lambda\, d\mu\, d\nu$, получим

$$\operatorname{div}\mathbf{A} = \frac{1}{l_\lambda l_\mu l_\nu}\left[\frac{\partial(l_\mu l_\nu A_\lambda)}{\partial\lambda} + \frac{\partial(l_\lambda l_\nu A_\mu)}{\partial\mu} + \frac{\partial(l_\lambda l_\mu A_\nu)}{\partial\nu}\right].$$

Чтобы получить выражение для ротора, можно исходить из его определения (67). Циркуляция вектора **A** по бесконечно малому прямоугольнику, перпендикулярному к вектору $\mathbf{e}_\lambda$, с точностью до малых высшего порядка, равна (рис. 333)

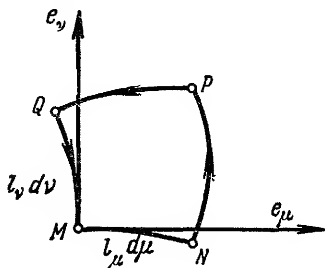


Рис. 333.

$$\oint \mathbf{A}\cdot d\mathbf{r} = \left(\int\limits_{NP} - \int\limits_{MQ}\right) - \left(\int\limits_{QP} - \int\limits_{MN}\right) =$$

$$= \partial_\mu (A_\nu\, ds_\nu) - \partial_\nu (A_\mu\, ds_\mu) =$$

$$= \partial_\mu (l_\nu A_\nu\, d\nu) - \partial_\nu (l_\mu A_\mu\, d\mu) =$$

$$= \left[\frac{\partial(l_\nu A_\nu)}{\partial\mu} - \frac{\partial(l_\mu A_\mu)}{\partial\nu}\right] d\mu\, d\nu.$$

Производя деление на элементарную площадку $dS = l_\mu l_\nu\, d\mu\, d\nu$, а затем циклическую перестановку индексов, получим

$$(\operatorname{rot}\mathbf{A})_\lambda = \frac{1}{l_\mu l_\nu}\left[\frac{\partial(l_\nu A_\nu)}{\partial\mu} - \frac{\partial(l_\mu A_\mu)}{\partial\nu}\right],$$

$$(\operatorname{rot}\mathbf{A})_\mu = \frac{1}{l_\lambda l_\nu}\left[\frac{\partial(l_\lambda A_\lambda)}{\partial\nu} - \frac{\partial(l_\nu A_\nu)}{\partial\lambda}\right],$$

$$(\operatorname{rot}\mathbf{A})_\nu = \frac{1}{l_\lambda l_\mu}\left[\frac{\partial(l_\mu A_\mu)}{\partial\lambda} - \frac{\partial(l_\lambda A_\lambda)}{\partial\mu}\right].$$

Для плоских полей все указанные формулы естественно упрощаются: надо положить $A_\nu = 0$, $l_\nu = 1$ и считать все участвующие величины не зависящими от ν.

29. Общая формула для преобразования интегралов. Оказывается, что формулы Стокса, Остроградского и аналогичные формулы в многомерном пространстве можно записать в виде единой формулы. Для этого допустим, что в k-мерном пространстве E_k с мерой Лебега (п. 20) дано ориентированное $(p+1)$-мерное $(p = 1, 2, \ldots, k-1)$ многообразие (Ω) с p-мерной границей (Ω'). Ориентация (Ω) порождает соответствующую ориентацию (Ω') по следующему правилу: если некоторый малый $(p+1)$-мерный тетраэдр $A_1A_2A_3\ldots A_{p+1}A_{p+2}$, вершины которого перечислены в соответствии с ориентацией (Ω), расположен так, что его грань $A_1A_2A_3\ldots A_{p+1}$ принадлежит (Ω'), то этот порядок вершин должен соответствовать ориентации (Ω'). (Проверьте, что если (Ω) — поверхность в трехмерном пространстве, то это правило совпадает с обычным правилом согласования ориентаций поверхности и ее контура.)

Пусть рассматривается интеграл вида (54), где $(S) = (\Omega')$. Подынтегральное выражение

$$\omega = \sum_{m_1, m_2, \ldots, m_p = 1}^{k} u_{m_1; \ldots; m_p}(t_1, \ldots, t_k)\, dt_{m_1} \ldots dt_{m_p}, \qquad (73)$$

однородное, измерения p относительно $dt_1, \ldots, dt_k$, называется *дифференциальной формой* измерения p.

Над дифференциальными формами можно производить некоторые действия. Так, формы одинакового измерения можно складывать; по существу, и выражение (73) представляет собой сумму простейших форм — одночленов. Формы можно перемножать, причем в соответствии с определением интеграла (53) принято считать, что при перестановке двух дифференциалов в одночленной форме она множится на -1, а при наличии двух одинаковых дифференциалов она равна нулю. Форму можно множить на константу или на функцию от $t_1, t_2, \ldots, t_k$; впрочем, последнюю можно рассматривать как форму нулевого измерения. При этом соблюдаются обычные правила сложения и умножения, однако умножение, вообще говоря, не перестановочно.

Дифференциальную форму можно дифференцировать по правилу

$$d\omega = d\left(\sum u_{m_1, m_2, \ldots, m_p}\, dt_{m_1} \ldots dt_{m_p}\right) = \sum du_{m_1, \ldots, m_p}\, dt_{m_1} \ldots dt_{m_p} = \\ = \sum \left(\frac{\partial u_{m_1, \ldots, m_p}}{\partial t_1} dt_1 + \ldots + \frac{\partial u_{m_1, \ldots, m_p}}{\partial t_k} dt_k\right) dt_{m_1} \ldots dt_{m_p},$$

где надо еще раскрыть скобки и привести подобные члены. Как видим, при дифференцировании формы ее измерение повышается на единицу.

Оказывается (мы не будем этого доказывать), что при таких определениях имеет место следующая *общая формула преобразования интеграла* (54):

$$\overbrace{\int \ldots \int}^{p \text{ раз}}_{(\Omega')} \omega = \overbrace{\iint \ldots \int}^{(p+1) \text{ раз}}_{(\Omega)} d\omega. \tag{74}$$

Рассмотрим, например, случай $k=2$, $p=1$. Если писать x, y вместо t_1, t_2 и обозначить

$$\omega = P(x, y)\, dx + Q(x, y)\, dy,$$

то

$$d\omega = dP\, dx + dQ\, dy = \\ = \left(\frac{\partial P}{\partial x} dx + \frac{\partial P}{\partial y} dy\right) dx + \left(\frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy\right) dy = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy,$$

откуда формула (74), с точностью до обозначений, приобретает вид формулы Грина (70). Мы предлагаем читателю разобрать случай $k=3$, $p=1$ (формула Стокса) и $k=3$, $p=2$ (формула Остроградского). При этом следует учесть выражение потока в виде двойного интеграла по координатам (ср. п. 22):

$$\int_{(\sigma)} \mathbf{A} \cdot d\boldsymbol{\sigma} = \int_{(\sigma)} \mathbf{A} \cdot \mathbf{n}\, d\sigma = \int_{(\sigma)} [A_x \cos(\widehat{\mathbf{n}, x}) + A_y \cos(\widehat{\mathbf{n}, y}) + A_z \cos(\widehat{\mathbf{n}, z})]\, d\sigma = \\ = \iint_{(\sigma)} (A_x\, dy\, dz + A_y\, dz\, dx + A_z\, dx\, dy).$$

РЯДЫ

Ряды уже встречались в нашем курсе, начиная с п. III.6, который мы советуем просмотреть перед чтением этой главы. В частности, в п. III.6 были даны основные определения *сходимости* и *суммы* числового ряда. Теперь мы переходим к систематическому изложению теории рядов.

§ 1. Числовые ряды

1. Ряды с положительными членами. Рассмотрим ряд

$$a_1+a_2+\ldots+a_n+\ldots, \quad \text{все} \quad a_n \geqslant 0; \tag{1}$$

обозначим, как и в п. III.6, частные суммы этого ряда через S_1, S_2, ..., S_n, ... В данном случае $S_1 \leqslant S_2 \leqslant \ldots \leqslant S_n \leqslant \ldots$ (почему?), и, вспомнив два способа изменения возрастающей величины (п. III.5), мы приходим к выводу, что для ряда (1) может быть один из двух случаев: он или сходится, или расходится к бесконечности. Это можно записать так:

$$\sum_{k=1}^{\infty} a_k < \infty \quad \text{или} \quad \sum_{k=1}^{\infty} a_k = \infty.$$

Подчеркнем, что первое неравенство как обозначение сходимости пишут только для рядов с положительными членами.

Если, кроме ряда (1), рассматривается ряд

$$b_1+b_2+\ldots+b_n+\ldots, \tag{2}$$

причем

$$0 \leqslant a_k \leqslant b_k \quad (k=1,\ 2,\ 3,\ \ldots), \tag{3}$$

то

$$\sum_{k=1}^{\infty} a_k \leqslant \sum_{k=1}^{\infty} b_k.$$

Действительно, это сразу следует из аналогичного неравенства для частных сумм. Отсюда вытекает *признак сравнения*, аналогичный признаку п. XIV.15: при выполнении условия (3), если ряд (2) сходится, то и ряд (1) сходится, а если ряд (1) расходится, то и ряд (2) расходится.

Например, ряд

$$\frac{1}{3^2 \ln 2}+\frac{1}{3^3 \ln 3}+\frac{1}{3^4 \ln 4}+\ldots$$

сходится, что следует из сравнения его с рядом (III.6):

$$\frac{1}{3^n \ln n}<\frac{1}{3^n} \quad (n=3,\ 4,\ \ldots);$$

правда, первые члены рядов не подчиняются этой оценке, но согласно п.III.6 это на факт сходимости не влияет.

С первым признаком сравнения связан другой признак: если

$$a_k>0, \quad b_k>0 \qquad (k=1,\ 2,\ \ldots),$$

$$\frac{a_k}{b_k}\underset{k\to\infty}{\longrightarrow}\mathrm{const}\neq 0, \quad \neq\infty,$$

то ряды (1) и (2) сходятся или расходятся одновременно. Действительно, из указанного условия вытекает, что отношение $\frac{a_k}{b_k}$ для всех k заключено между некоторыми *положительными* постоянными m и M:

$$m\leqslant\frac{a_k}{b_k}\leqslant M, \quad \text{т. е.} \quad mb_k\leqslant a_k\leqslant Mb_k.$$

Отсюда, суммируя по k от 1 до n, а затем переходя к пределу при $n\to\infty$, получаем

$$m\sum_{k=1}^{\infty} b_k\leqslant\sum_{k=1}^{\infty} a_k\leqslant M\sum_{k=1}^{\infty} b_k,$$

откуда и вытекает наше утверждение (почему?).

Широко применяется следующий *достаточный признак Даламбера*: *если для ряда* (1) *существует предел*

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=l,$$

то в случае $l<1$ *ряд* (1) *сходится, а в случае* $l>1$ — *расходится.* Второе получается совсем просто, так как при $l>1$ отношение $\frac{a_{n+1}}{a_n}$, приближающееся с возрастанием n к l, становится, начиная с некоторого номера, бо́льшим единицы, т. е. члены ряда для больших n возрастают, а значит, не выполнен необходимый признак сходимости (п. III.6). Пусть теперь $l<1$; выберем какое-нибудь постоянное число l' между l и 1. Так как отношение $\frac{a_{n+1}}{a_n}$ безгранично приближается к l, то оно обязательно, начиная с некоторого номера N, станет меньше l'. Но тогда

$$\frac{a_{N+1}}{a_N}<l', \quad \frac{a_{N+2}}{a_{N+1}}<l', \quad \frac{a_{N+3}}{a_{N+2}}<l', \ \ldots,$$

откуда

$a_{N+1} < a_N l',\quad a_{N+2} < a_{N+1} l' < a_N l'^2,\quad a_{N+3} < a_{N+2} l' < a_N l'^3$ и т. д.,

т. е. члены ряда (1), начиная с номера N, меньше членов ряда

$$a_N + a_N l' + a_N l'^2 + a_N l'^3 + \ldots$$

Так как $0 < l' < 1$, то члены последнего ряда образуют бесконечную убывающую геометрическую прогрессию и потому он сходится (см. ряд (III 7)), а отсюда по признаку сравнения сходится и ряд (1).

Рассмотрим, например, ряд

$$\sum_{n=1}^{\infty} \frac{a^n}{n^p} \quad (a > 0,\ p \gtrless 0). \tag{4}$$

Для применения признака Даламбера надо найти предел

$$\lim_{n\to\infty} \left(\frac{a^{n+1}}{(n+1)^p} : \frac{a^n}{n^p}\right) = \lim_{n\to\infty} \frac{a}{\left(1+\frac{1}{n}\right)^p} = a.$$

Значит, ряд (4) сходится при $a < 1$ и расходится при $a > 1$, при $a = 1$ признак Даламбера ответа не дает.

Когда признак Даламбера отказывает, часто бывает возможно применить более сильный *достаточный интегральный признак Коши*: если выражение для a_n имеет смысл не только для целых $n = 1, 2, 3, \ldots$, но и для всех действительных $n \geqslant 1$ и убывает с ростом n, то

$$\int_1^{\infty} a_n\,dn \leqslant \sum_{n=1}^{\infty} a_n \leqslant \int_1^{\infty} a_n\,dn + a_1; \tag{5}$$

Рис 334.

значит, ряд (1) сходится или расходится одновременно с выписанным интегралом. Для доказательства заметим, что на основании рис. 334, *а* будет

$$\int_1^N a_n\,dn \leqslant a_1\cdot 1 + a_2\cdot 1 + \ldots + a_N\cdot 1, \tag{6}$$

а на основании рис. 334, *б*

$$\int_1^N a_n\,dn \geqslant a_2\cdot 1 + a_3\cdot 1 + \ldots + a_N\cdot 1,$$

откуда

$$a_1 + a_2 + a_3 + \ldots + a_N \leqslant \int_1^N a_n\,dn + a_1. \qquad (7)$$

Если в неравенствах (6) и (7) перейти к пределу при $N \to \infty$, то получим (5).

Рассмотрим, например, ряд

$$\sum_{n=1}^{\infty} \frac{1}{n^p}, \qquad (8)$$

который получается из ряда (4) при $a=1$, когда признак Даламбера не действует. Так как $a_n = n^{-p}$ при $p>0$ удовлетворяет условиям признака Коши, то ряд (8) сходится или расходится одновременно с интегралом

$$\int_1^{\infty} \frac{1}{n^p}\,dn.$$

Но непосредственное вычисление последнего интеграла показывает, что он сходится только при $p>1$ (см. вычисление интеграла (XIV.51)). Значит, и ряд (8) сходится только при $p>1$. В частности, при $p=1$ получаем так называемый *гармонический ряд*

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots = \infty.$$

Формулы (6) и (7) можно применить для двусторонней оценки частной суммы расходящегося ряда, откуда можно получить асимптотическую формулу для такой суммы в зависимости от номера. Аналогично проводится оценка любой суммы большого числа слагаемых, монотонно зависящих от номера. Для уточнения результата можно несколько наибольших слагаемых просуммировать непосредственно, а оценивать лишь оставшиеся, так как тогда разность между оценками сверху и снизу сблизится.

Более точные приближения, хотя и без двусторонних оценок, получаются с помощью формул численного интегрирования (п. XIV.13). Покажем, например, применение формулы Симпсона к приближенному вычислению суммы

$$S_{m,N} = \frac{1}{m} + \frac{1}{m+1} + \ldots + \frac{1}{N} \qquad (m=1,\ 2,\ \ldots;\ N \geqslant m+2).$$

Для этого пишем на основе формулы (XIV.39) при $h=1$

$$\int_k^{k+1} \frac{1}{x}\,dx + \int_{k+1}^{k+2} \frac{1}{x}\,dx = \int_k^{k+2} \frac{1}{x}\,dx \approx \frac{1}{3}\left(\frac{1}{k} + \frac{4}{k+1} + \frac{1}{k+2}\right).$$

Суммируя эти формулы при $k=m,\ m+1,\ \ldots,\ N-1$, после несложных преобразований и интегрирования получаем

$$\ln N - \ln m + \ln(N+1) - \ln(m+1) \approx \frac{1}{3}\left(6S_{m,N} - \frac{5}{m} - \frac{1}{m+1} - \frac{1}{N} + \frac{1}{N+1}\right),$$

откуда

$$S_{m,N} \approx \ln N + \frac{6m+5}{6m(m+1)} - \frac{1}{2}\ln(m^2+m) + \frac{1}{2}\ln\left(1+\frac{1}{N}\right) + \frac{1}{6N(N+1)}. \quad (9)$$

Рассматривая рис. 334, *а* для функции $a_n = \frac{1}{n}$, легко проверить, что существует конечный положительный предел

$$Э = \lim_{N\to\infty}\left[\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{N}-\int_1^N \frac{1}{n}\,dn\right] = \lim_{N\to\infty}(S_{1,N}-\ln N);$$

он называется *постоянной Эйлера* Из равенства (9) и формулы $S_{m,N} = S_{1,N} - \frac{1}{1} - \frac{1}{2} - \ldots - \frac{1}{m-1}$ вытекает приближенное значение постоянной Эйлера

$$Э \approx \frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{m-1}+\frac{6m+5}{6m(m+1)}-\frac{1}{2}\ln(m^2+m),$$

тем более точное, чем больше m. Так, при $m=1$ и $m=2$ получаем соответственно значения 0,570 и 0,576, тогда как $Э=0{,}577$, с точностью до 0,001.

По поводу вычисления сумм при помощи интегралов см. [21], §§ I.2 и III.4.

2. Ряды с членами любого знака. Перейдем теперь к рядам

$$a_1 + a_2 + \ldots + a_n + \ldots, \quad a_n \gtrless 0 \quad (10)$$

с членами произвольного знака. Здесь, прежде всего, имеет место утверждение: *если*

$$\sum_{k=1}^{\infty} |a_k| < \infty, \quad (11)$$

то и ряд (10) *сходится*; в этом случае он называется *абсолютно сходящимся*. Доказательство этого совершенно аналогично доказательству подобного свойства в п. XIV.15, и мы его предоставим желающим. Если ряд (11) расходится, то ряд (10) может все же сходиться; это — *неабсолютная сходимость*.

Чаще всего, применяя признаки п. 1, проверяют именно абсолютную сходимость ряда (10). Например, если

$$\lim_{n\to\infty}\frac{|a_{n+1}|}{|a_n|} < 1,$$

то по признаку Даламбера ряд (11) сходится, а потому ряд (10) абсолютно сходится (если этот предел больше единицы, то не выполнен необходимый признак сходимости и ряд расходится) и т. п.

Следующий *достаточный признак Лейбница* гарантирует сходимость, но не обязательно абсолютную: если для ряда

$$a_1 - a_2 + a_3 - a_4 + \ldots \quad (12)$$

будет $a_1 > a_2 > a_3 > \ldots \to 0$, то он сходится. Для доказательства заметим, что если изображать частные суммы ряда (12) на некоторой оси S (рис. 335), то каждый из следующих переходов от 0 к S_1, от S_1 к S_2, от S_2 к S_3 и т. д. будет совершаться в направлении, противоположном предыдущему, причем на меньшее расстояние; другими словами, в данном случае $0 < S_2 < S_1$, $S_2 < S_3 < S_1$, $S_2 < S_4 < S_3$, ... (рис. 335). Значит, четные частные суммы (т. е. частные суммы с четными номерами) образуют возрастающую ограниченную последовательность и потому (п. III.5, свойство 10) имеют предел S', а нечетные частные суммы образуют убывающую ограниченную последовательность и потому имеют предел S''. Переходя в равенстве $S_{2n+1} = S_{2n} + a_{2n+1}$ к пределу при $n \to \infty$, получим, что $S'' = S'$, т. е. все частные суммы имеют единый предел и тем самым ряд (12) сходится. Заодно мы видим, что *сумма ряда* (12) *заключена между любой четной и любой нечетной частными суммами*, что дает возможность произвести оценку суммы ряда.

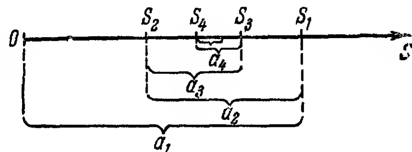


Рис. 335.

Например, ряд

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^p}$$

при $p > 0$ удовлетворяет условиям признака Лейбница и потому сходится. В то же время эта сходимость будет абсолютной только при $p > 1$ (см. п. 1).

Отметим в заключение, что из неравенства

$$\left| \sum_{k=1}^{n} a_k \right| \leqslant \sum_{k=1}^{n} |a_k|$$

после перехода к пределу при $n \to \infty$ для сходящегося ряда (10) получается неравенство

$$\left| \sum_{k=1}^{\infty} a_k \right| \leqslant \sum_{k=1}^{\infty} |a_k|.$$

3. Действия с рядами. 1. *Сходящиеся ряды можно почленно складывать*, т. е. если

$$a_1 + a_2 + \ldots + a_n + \ldots = S, \quad b_1 + b_2 + \ldots + b_n + \ldots = T,$$

то

$$(a_1 \pm b_1) + (a_2 \pm b_2) + \ldots + (a_n \pm b_n) + \ldots = S \pm T.$$

Для доказательства надо заметить, что частная сумма последнего ряда $P_n = S_n \pm T_n$, а затем перейти к пределу при $n \to \infty$.

На основе этого свойства иногда производится следующее преобразование: ряд с членами любого знака, например

$$a-b-c+d+e+f-g+\ldots=S \tag{13}$$

(где все a, b, c, ... считаются >0), представляется в виде разности рядов с неотрицательными членами

$$\left.\begin{aligned} &a+0+0+d+e+f+0+\ldots=S_1 \\ \text{и} \quad &0+b+c+0+0+0+g+\ldots=S_2, \end{aligned}\right\} \tag{14}$$

т. е. суммируются все положительные и все отрицательные члены, а результаты вычитаются один из другого. Так можно поступать только с абсолютно сходящимся рядом (13), так как для неабсолютно сходящегося ряда (13) оба ряда (14) имеют бесконечную сумму (почему?). В последнем случае сходимость ряда (13) получается за счет «баланса» между этими бесконечностями, в результате которого частные суммы обоих рядов (14) нарастают с одинаковой скоростью.

Аналогично первому свойству проверяется свойство

2. *Сходящийся ряд можно почленно множить на общий множитель*: если

$$a_1+a_2+\ldots+a_n+\ldots=S,$$

то

$$ka_1+ka_2+\ldots+ka_n+\ldots=kS.$$

3. *В сходящемся ряде можно произвольно группировать члены*: например, если

$$a_1+a_2+a_3+a_4+a_5+a_6+a_7+a_8+\ldots=S, \tag{15}$$

то

$$(a_1+a_2)+a_3+(a_4+a_5+a_6)+(a_7+a_8)+\ldots=S. \tag{16}$$

Действительно, если частные суммы первого ряда S_1, S_2, S_3, ... $\to S$, то частные суммы второго ряда, равные S_2, S_3, S_6, S_8, ..., также стремятся к S.

Если ряд (15) расходится к бесконечности, то аналогично получаем, что и ряд (16) расходится к бесконечности. Если же ряд (15) расходится колебательным образом (п. III.6), то ряд (16) может расходиться, а может и сходиться, причем сумма тогда будет зависеть от способа группировки, т. е. расстановки скобок. Например, для ряда (III.9) имеем

$$(1-1)+(1-1)+(1-1)+\ldots=0+0+0+\ldots=0,$$
$$1+(-1+1)+(-1+1)+\ldots=1+0+0+\ldots=1.$$

Пока не была осознана разница между сходящимися и расходящимися рядами, этот факт воспринимался как необъяснимый парадокс. Современное определение понятия суммы сходящегося ряда стало возможным только на базе развитой теории пределов и было дано Коши в 1821 г., хотя ряды широко применялись и в XVII—XVIII вв.

4. *В ряде с неотрицательными членами можно произвольно переставлять порядок членов (слагаемых), отчего сумма не изменится.*

Дело в том, что если составлять последовательные частные суммы для ряда с переставленными (без пропусков) членами, то любой член исходного ряда будет входить в эти суммы, начиная с некоторой. Поэтому и любая частная сумма исходного ряда составит часть частной суммы с достаточно большим номером переставленного ряда и потому не превосходит всей суммы переставленного ряда. Но тогда и предел частных сумм исходного ряда, т. е. сумма исходного ряда, не превосходит суммы переставленного ряда. А так как первый ряд получается из второго также перестановкой членов, т. е. и сумма второго ряда не может превосходить суммы первого, то обе суммы равны.

В абсолютно сходящемся ряде также можно произвольно переставлять порядок членов.

Действительно, как было указано в свойстве 1, такой ряд можно представить как разность двух сходящихся рядов с неотрицательными членами, так что перестановка членов у исходного ряда сводится к перестановке членов у этих двух рядов, что, как мы только что видели, не меняет их сумм.

Перестановка членов в неабсолютно сходящемся ряде может изменить сумму или даже сделать ряд расходящимся. Дело в том, что в результате такой перестановки можно изменить относительную скорость нарастания рядов (14), что и приводит к этому на первый взгляд парадоксальному результату, который можно продемонстрировать на следующем примере.

Сумма S ряда

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\frac{1}{9}-\frac{1}{10}+\frac{1}{11}-\ldots=S \qquad (17)$$

согласно п. 2 заключена между $S_1=1$ и $S_2=0,5$. Отсюда в силу свойства 2

$$\frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\frac{1}{10}-\ldots=\frac{S}{2},$$

а потому и

$$0+\frac{1}{2}+0-\frac{1}{4}+0+\frac{1}{6}+0-\frac{1}{8}+0+\frac{1}{10}+0-\frac{1}{12}+\ldots=\frac{S}{2}.$$

Произведя почленное сложение этого ряда с рядом (17), получим

$$1+0+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+0+\frac{1}{7}-\frac{1}{4}+\frac{1}{9}+0+\frac{1}{11}-\frac{1}{6}+\ldots=\frac{3}{2}S,$$

т. е.

$$1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\frac{1}{9}+\frac{1}{11}-\frac{1}{6}+\ldots=\frac{3}{2}S.$$

Однако последний ряд получается из ряда (17) перестановкой членов (проверьте!), а сумма, как видим, изменилась.

Таким образом, в процессах, так или иначе связанных с перестановкой членов ряда, только с абсолютно сходящимися рядами можно смело обращаться так же, как с конечными суммами; при действиях же с неабсолютно сходящимися рядами надо соблюдать определенную осторожность.

4. Скорость сходимости ряда. Для практического подсчета суммы ряда обычно вычисляют частную сумму нескольких его первых членов, а остальные просто отбрасывают, если имеются основания полагать, что они не повлияют существенно на значение суммы (ср. вычисление числа e в п. IV 16). При этом *нужно, чтобы ряд не просто сходился, но быстро сходился*, т. е. чтобы, взяв небольшое число членов, мы почти исчерпали бы полную сумму, получив ее с хорошей точностью. Если же ряд сходится медленно, плохо, то для непосредственных практических вычислений он бывает часто

непригоден, хотя из него бывает возможно получить другие, быстро сходящиеся ряды; впрочем, иногда остаток такого ряда удается приближенно выразить с помощью интегралов, наподобие п. 1. Очень медленно сходятся неабсолютно сходящиеся ряды (п. 2); но и абсолютно сходящиеся ряды часто сходятся медленно.

Скорость сходимости ряда определяется в основном скоростью стремления его общего члена к нулю с возрастанием номера. Часто сравнительно плохо сходятся ряды, для которых общий член a_n имеет порядок n^{-p} (т. е. $a_n = O(n^{-p})$; см. п III.11) при $p > 1$, причем сходимость тем лучше, чем больше p. Лучше сходятся ряды, для которых a_n имеет порядок q^n при $0 < q < 1$, это скорость геометрической прогрессии (III.7), причем сходимость тем лучше, чем меньше q. Еще лучше сходятся ряды, для которых a_n имеет порядок $\frac{1}{n!}$ и т. д.

Впрочем, это лишь самые общие установки, так как при любом порядке общего члена для $n \to \infty$ в данном вопросе может оказаться существенным поведение первых членов ряда.

Если ряд

$$a_1 + a_2 + \ldots + a_n + \ldots \tag{18}$$

сходится медленно, то часто стремятся перейти к более быстро сходящемуся ряду. Один из способов этого такой. Подбирают ряд

$$b_1 + b_2 + \ldots + b_n + \ldots = \sigma$$

с известной суммой так, чтобы $a_n \sim b_n$ при $n \to \infty$ (п. III.7—8). Тогда $a_n = b_n + \gamma_n$, где $|\gamma_n| \ll |a_n|$, и потому ряд (18) можно представить в виде

$$(b_1 + \gamma_1) + (b_2 + \gamma_2) + \ldots =$$
$$= (b_1 + b_2 + \ldots) + (\gamma_1 + \gamma_2 + \ldots) = \sigma + \gamma_1 + \gamma_2 + \ldots + \gamma_n + \ldots,$$

а у этого ряда общий член стремится к нулю быстрее.

Для применения этого *способа выделения особенности* надо иметь набор рядов с известными суммами. Обычно пользуются геометрической прогрессией (III.7), рядами, указанными в п. IV.16, и их комбинациями, а также рядом

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = \zeta(p) \qquad (p > 1). \tag{19}$$

Последняя сумма, зависящая от p, называется *дзёта-функцией Римана* по имени выдающегося немецкого математика Б. Ри́мана (1826—1866) хотя ее впервые ввел Эйлер в 1737 г.; таблицы значений этой функции имеются, в частности, в книге [53].

Рассмотрим, например, ряд

$$S = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n^3+1}}. \tag{20}$$

Его члены эквивалентны при $n \to \infty$ членам ряда (19) при $p = \frac{3}{2}$, т. е. ряд (20), хотя и сходится, но очень медленно. С помощью оценки (5) легко проверить, что остаток ряда (19) эквивалентен $\frac{1}{(p-1)\,n^{p-1}}$, т. е. остаток ряда (20) имеет порядок $2n^{-1/2}$, и чтобы получить S с точностью до 0,01, надо взять около 40 000 членов! Но если воспользоваться методом выделения

особенности, получим

$$\frac{1}{\sqrt{n^3+1}}=\frac{1}{\sqrt{n^3}}+\gamma_n,$$

$$\gamma_n=\frac{\sqrt{n^3}-\sqrt{n^3+1}}{\sqrt{n^3+1}\sqrt{n^3}}=-\frac{1}{\sqrt{n^3}\sqrt{n^3+1}\,(\sqrt{n^3}+\sqrt{n^3+1})},$$

а потому ряд (20) можно представить в виде

$$S=\zeta\left(\frac{3}{2}\right)-\sum_{n=1}^{\infty}\frac{1}{\sqrt{n^3}\sqrt{n^3+1}\,(\sqrt{n^3}+\sqrt{n^3+1})}. \qquad (21)$$

По таблице первое слагаемое равно 2,612, а общий член последнего ряда эквивалентен $(2n^{9/2})^{-1}$, а потому его остаток эквивалентен $(7n^{7/2})^{-1}$, т. е. для точности до 0,01 надо взять три члена! Если S требуется с большей точностью, то из ряда (21) надо вновь выделить особенность, что даст (проверьте!)

$$S=\zeta\left(\frac{3}{2}\right)-\frac{1}{2}\zeta\left(\frac{9}{2}\right)+$$
$$+\sum_{n=1}^{\infty}\frac{24n^9+7n^6-2n^3-1}{2\sqrt{n^9}\,A\,(\sqrt{n^3}+A)\,(3n^3+1+\sqrt{n^3}\,A)\,(2\sqrt{n^3}A^3+2n^6-3n^3-1)},$$

где для краткости положили $\sqrt{n^3+1}=A$. Выделенная часть по таблице равна 2,085, а остаток последнего ряда асимптотически равен (т. е. эквивалентен) $\frac{3}{52}n^{-\frac{13}{2}}$. Значит, для вычисления S с точностью до 0,001 достаточно взять всего два-три члена.

Это последовательное выделение особенностей можно усовершенствовать, если воспользоваться рядом Тейлора (IV.60) для бинома:

$$\frac{1}{\sqrt{n^3+1}}=(n^3+1)^{-1/2}=n^{-3/2}\left(1+\frac{1}{n^3}\right)^{-1/2}=$$
$$=n^{-3/2}\left(1-\frac{1}{2n^3}+\frac{3}{8n^6}-\frac{5}{16n^9}+\ldots\right). \qquad (22)$$

Обрывая этот ряд на любом члене, например на третьем, получим

$$\frac{1}{\sqrt{n^3+1}}=n^{-3/2}\left(1-\frac{1}{2n^3}+\frac{3}{8n^6}\right)-\gamma_n, \qquad (23)$$

откуда

$$S=\zeta\left(\frac{3}{2}\right)-\frac{1}{2}\zeta\left(\frac{9}{2}\right)+\frac{3}{8}\zeta\left(\frac{15}{2}\right)-\sum_{n=1}^{\infty}\gamma_n=2,462-\sum_{n=1}^{\infty}\gamma_n,$$

причем γ_n, которые можно выразить точно с помощью (23), асимптотически равны $\frac{5}{16}n^{-21/2}$, на основе (22).

Аналогично преобразуются и другие ряды, причем наряду с (19) применяются ряды

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^p}=\sum_{n=1}^{\infty}\frac{1}{n^p}-\sum_{n=1}^{\infty}\frac{2}{(2n)^p}=\zeta(p)\left(1-\frac{1}{2^{p-1}}\right), \tag{24}$$

$$\sum_{n=1}^{\infty}\frac{1}{n(n+1)}=\sum_{n=1}^{\infty}\left(\frac{1}{n}-\frac{1}{n+1}\right)=$$
$$=\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\ldots=1,$$

$$\sum_{n=1}^{\infty}\frac{1}{n(n+1)(n+2)}=\frac{1}{2}\sum_{n=1}^{\infty}\left[\frac{1}{n(n+1)}-\frac{1}{(n+1)(n+2)}\right]=$$
$$=\frac{1}{2}\left[\left(\frac{1}{1\cdot 2}-\frac{1}{2\cdot 3}\right)+\left(\frac{1}{2\cdot 3}-\frac{1}{3\cdot 4}\right)+\ldots\right]=\frac{1}{4}\ \text{и т. д.}$$

Формула (24) справедлива и при $0<p<1$, причем $\zeta(p)$ для таких p определена не по формуле (19), так как ряд расходится, а иным способом, о котором мы здесь не будем говорить. При $p=1$, как вытекает из формулы (IV.61), левая часть ряда (24) равна $\ln 2$.

Для знакочередующихся рядов (п. 2) имеется большая опасность *потери точности*. Рассмотрим для примера степенной ряд (IV.56) для косинуса при $x=100$:

$$\cos 100=1-\frac{100^2}{2!}+\frac{100^4}{4!}-\frac{100^6}{6!}+\frac{100^8}{8!}-\ldots \tag{25}$$

Ряд, стоящий в правой части, сходится и даже абсолютно (почему?), но практическое его использование невозможно. Действительно, хотя члены, начиная с 51-го, убывают, причем, в конце концов, достаточно быстро, что существенно для теоретической сходимости, но до этого они успевают кошмарно возрасти. А так как вся сумма должна получиться по абсолютной величине меньше единицы, то все эти члены «почти» взаимно уничтожаются. Как известно (п. I.9), такая ситуация очень опасна, так как все вычисления пришлось бы проводить с огромным числом значащих цифр и объем вычислительной работы был бы неоправданно велик. Поэтому рядов типа (25) надо избегать, а если они появляются, надо преобразовывать их в другие ряды, поддающиеся практическому вычислению; так, в данном примере можно воспользоваться периодичностью косинуса и перейти к значительно меньшему аргументу.

5. Ряды с комплексными, векторными и матричными членами. Определение сходимости и суммы ряда с комплексными членами

$$z_1+z_2+\ldots+z_n+\ldots,\quad z_n=x_n+iy_n\qquad (n=1,\ 2,\ 3,\ \ldots) \tag{26}$$

дается в точности так, как для ряда с вещественными членами (п. III.6). При этом иногда разбивают ряд (26) на два:

$$x_1+x_2+\ldots+x_n+\ldots,\quad y_1+y_2+\ldots+y_n+\ldots \tag{27}$$

Если оба ряда (27) сходятся и имеют сумму соответственно x и y, то и ряд (26) сходится и имеет сумму $z=x+iy$. Если хоть один

из рядов (27) расходится, то и ряд (26) расходится. Так как ряды (27) вещественные, то к ним применимы методы п. 2.

Применяется также следующий простой признак: если

$$\sum_{n=1}^{\infty} |z_n| < \infty, \tag{28}$$

то оба ряда (27) абсолютно сходятся, а потому ряд (26) также сходится; в этом случае ряд (26) называется *абсолютно сходящимся*. К ряду (28) можно применять методы п. 1.

Аналогично рассматриваются ряды с векторными членами

$$\mathbf{u}_1 + \mathbf{u}_2 + \ldots + \mathbf{u}_n + \ldots \tag{29}$$

Если все $\mathbf{u}_n$ — векторы в трехмерном пространстве x, y, z, то для перехода к скалярным рядам надо произвести проектирование ряда (29) на все три оси x, y, z.

Рассматриваются также ряды вида

$$\mathbf{A}_1 + \mathbf{A}_2 + \ldots + \mathbf{A}_n + \ldots, \tag{30}$$

членами которого служат матрицы (п. XI.2) одного и того же размера. Для сходимости ряда (30) необходимо и достаточно, чтобы сходился каждый из рядов, состоящих из соответственных, т. е. одинаково расположенных элементов этих матриц.

Свойства рядов (26), (29) и (30) те же, что и вещественных рядов (п. 3).

6. Кратные ряды. Конечные суммы могут иметь не только один, но и два и более индексов суммирования.

Например,

$$\sum_{i=1}^{2} \sum_{j=1}^{3} a_{ij} = a_{11} + a_{12} + a_{13} + a_{21} + a_{22} + a_{23},$$

$$\sum_{i=1}^{4} \sum_{j=1}^{i} \frac{1}{i^j} = \frac{1}{1^1} + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{3^1} + \frac{1}{3^2} + \frac{1}{3^3} + \frac{1}{4^1} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4}$$

и т. п. (ср. п. XVI.8).

Бесконечные ряды также могут иметь более одного индекса суммирования; такие ряды называются *двойными*, *тройными* и т. д., вообще *кратными*. Мы рассмотрим только двойной ряд простейшего вида

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}; \tag{31}$$

рассмотрение рядов высшей кратности, а также рядов, у которых пределы суммирования во внутренней сумме переменные, как в приведенном выше примере, проводится аналогично.

Пусть сначала все $a_{ij} \geqslant 0$. Расположим все члены ряда (31) в простой, однократный ряд в каком-либо порядке, например

$$a_{11} + a_{12} + a_{21} + a_{13} + a_{22} + a_{31} + a_{14} + a_{23} + a_{32} + a_{41} + \ldots \tag{32}$$

Сумма этого ряда, не зависящая от порядка слагаемых (свойство 4 п. 3), и называется суммой ряда (31). При этом могут быть два случая, сходимости и расходимости, т. е.

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} < \infty \quad \text{или} \quad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} = \infty.$$

Таким образом, при $a_{ij} \geqslant 0$ сумма ряда (31) не зависит от способа суммирования, лишь бы при этом ни один член не был забыт. В частности, можно суммировать так:

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} = \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} a_{ij}\right) = \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} a_{ij}\right). \tag{33}$$

Если члены a_{ij} любого знака или даже комплексные, то наиболее просто рассматривается случай *абсолютной сходимости*

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |a_{ij}| < \infty. \tag{34}$$

Тогда и ряд (31) сходится, и для подсчета его суммы можно воспользоваться любой из формул вида (32) или (33), формулой

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} = \lim_{M,\, N \to \infty} \sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}$$

и т. п. Если условие (34) нарушено, то при суммировании ряда (31) может играть роль порядок суммирования (п. 3) и положение значительно осложняется.

Двойной ряд получается, в частности, при *перемножении двух абсолютно сходящихся рядов*

$$S_1 = \sum_{i=1}^{\infty} a_i \quad \text{и} \quad S_2 = \sum_{i=1}^{\infty} b_i = \sum_{j=1}^{\infty} b_j;$$

перед умножением мы изменили обозначение индекса суммирования в одном из рядов. Умножение производится так:

$$S_1 S_2 = \sum_{i=1}^{\infty} a_i \sum_{j=1}^{\infty} b_j = \sum_{i=1}^{\infty}\left(a_i \sum_{j=1}^{\infty} b_j\right) = \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} a_i b_j\right) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_i b_j;$$

при этом абсолютная сходимость использована в последнем равенстве, так как применена формула (33). Итак, умножение таких рядов производится по тому же правилу, что и умножение конечных сумм (каждый член первого ряда умножается на каждый член второго ряда), после чего получается абсолютно сходящийся двойной ряд. Аналогичный результат получится, если множителей больше.

§ 2. Общие функциональные ряды

7. Уклонение функций. Если членами ряда являются не числа, как в § 1, а функции, то возникает вопрос, в каком смысле частные суммы, которые являются функциями, сходятся (приближаются) к полной сумме, которая также является функцией, т. е. как оценить отличие двух функций друг от друга? Оказывается, это можно делать различными, неравносильными способами, тогда как отличие двух чисел a и b друг от друга всегда оценивается величиной $|a-b|$.

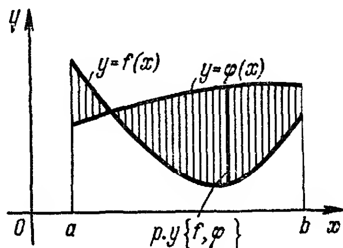


Рис. 336.

Пусть две функции, $f(x)$ и $\varphi(x)$, заданы на одном и том же конечном интервале $a \leqslant x \leqslant b$. *Равномерным* или *чебышевским уклонением* их друг от друга называется величина

$$\text{р. у.}\{f, \varphi\} = \max_{a \leqslant x \leqslant b} |f(x) - \varphi(x)|, \quad (35)$$

геометрический смысл которой показан на рис. 336. Это уклонение применяется только для ограниченных функций, причем, как правило, для непрерывных. Если равномерное уклонение двух функций друг от друга мало, то в каждой точке интервала $a \leqslant x \leqslant b$ значения $f(x)$ и $\varphi(x)$ мало отличаются друг от друга, и наоборот.

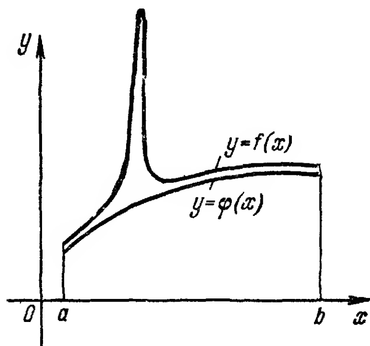


Рис. 337.

Средним интегральным уклонением функций $f(x)$ и $\varphi(x)$ друг от друга называется величина

$$\text{с. и. у.}\{f, \varphi\} = \int_a^b |f(x) - \varphi(x)|\, dx; \quad (36)$$

ее геометрический смысл — это площадь, заштрихованная на рис. 336, без учета знаков Для вычислений более удобно *среднее квадратичное уклонение*

$$\text{с. к. у.}\{f, \varphi\} = \sqrt{\int_a^b [f(x) - \varphi(x)]^2\, dx}, \quad (37)$$

которое во многом аналогично с. и. у. Эти уклонения применяются не только для непрерывных, но и для разрывных функций, и даже для неограниченных функций, если интеграл, который тогда будет несобственным, сходится (п. XIV.16). Имеются и другие виды уклонений.

Если в формулах (36) и (37) заменить подынтегральную функцию ее максимальным значением (35), то от этого интеграл может только увеличиться, и мы получим

$$\text{с. и. у.}\{f, \varphi\} \leqslant (b-a)\cdot\text{р. у.}\{f, \varphi\}; \quad \text{с. к. у.}\{f, \varphi\} \leqslant \sqrt{b-a}\cdot\text{р. у.}(f, \varphi). \quad (38)$$

Значит, если равномерное уклонение двух функций друг от друга мало́, то и среднее интегральное и среднее квадратичное уклонение тоже малы. Однако может так получиться, что равномерное уклонение двух функций велико, тогда как среднее — мало́ (рис. 337).

8. Сходимость функционального ряда. Рассмотрим ряд

$$f_1(x)+f_2(x)+\ldots+f_n(x)+\ldots, \tag{39}$$

членами которого являются функции, заданные на одном и том же конечном интервале $a\leqslant x\leqslant b$. Говорят, *что этот ряд сходится на данном интервале к функции $S(x)$ — сумме ряда, если уклонение частной суммы $S_n(x)=\sum_{k=1}^{n} f_k(x)$ от $S(x)$ стремится к нулю при возрастании n.* В зависимости от вида уклонения (п. 7) получается вид сходимости ряда (39). Так, ряд (39) сходится к сумме $S(x)$ *равномерно*, если

$$\max_{a\leqslant x\leqslant b} |S(x)-S_n(x)| \xrightarrow[n\to\infty]{} 0.$$

Подобным образом этот ряд сходится к $S(x)$ *в среднем* или *в среднем квадратичном* в зависимости от того, будет ли

$$\int_a^b |S(x)-S_n(x)|\,dx \xrightarrow[n\to\infty]{} 0$$

или

$$\sqrt{\int_a^b [S(x)-S_n(x)]^2\,dx} \xrightarrow[n\to\infty]{} 0.$$

Из оценок (38) видно, что *если ряд* (39) *сходится равномерно, то он сходится* и *в среднем и в среднем квадратичном* и притом к той же сумме. *Обратное не обязательно.*

Если ряд (39) на интервале $a\leqslant x\leqslant b$ *равномерно* сходится к сумме $S(x)$, то для *любого* числа c из этого интервала будет

$$f_1(c)+f_2(c)+\ldots+f_n(c)+\ldots=S(c);$$

в самом деле, отличие n-й частной суммы этого ряда от $S(c)$ не превосходит равномерного уклонения $S_n(x)$ и $S(x)$ (почему?) и тем самым стремится к нулю при $n\to\infty$. Это свойство дает возможность получить из функционального ряда с известной суммой числовые ряды с известными суммами.

Для проверки равномерной сходимости чаще всего применяется следующий *достаточный признак*, предложенный Вейерштрассом: *если все $|f_n(x)|\leqslant a_n$ $(n=1, 2, 3, \ldots;\ a\leqslant x\leqslant b)$, причем*

$$\sum_{n=1}^{\infty} a_n<\infty, \tag{40}$$

то ряд (39) *равномерно сходится.* Для доказательства заметим, что в данных условиях по признаку сравнения (п. 1) ряд (39) при каж-

дом зафиксированном x абсолютно сходится к сумме $S(x)$. При этом

$$\max_{a \leqslant x \leqslant b} |S(x) - S_n(x)| = \max_{a \leqslant x \leqslant b} \left| \sum_{k=n+1}^{\infty} f_k(x) \right| \leqslant$$
$$\leqslant \max_{a \leqslant x \leqslant b} \sum_{k=n+1}^{\infty} |f_k(x)| \leqslant \sum_{k=n+1}^{\infty} a_k,$$

а последняя сумма, как остаток сходящегося ряда (п. III.6), стремится к нулю с возрастанием номера.

Условие (40) иначе можно записать так:

$$\sum_{n=1}^{\infty} \max_{a \leqslant x \leqslant b} |f_n(x)| < \infty,$$

так как члены этого ряда можно обозначить через a_n. Аналогичный вид имеют достаточные признаки для сходимости ряда (39) в среднем и в среднем квадратичном:

$$\sum_{n=1}^{\infty} \int_a^b |f_n(x)|\,dx < \infty \text{ и соответственно } \sum_{n=1}^{\infty} \sqrt{\int_a^b [f_n(x)]^2\,dx} < \infty.$$

Мы не станем их доказывать.

Иногда бывает, что ряд (39) на интервале $a \leqslant x \leqslant b$ является расходящимся, но этот ряд сходится на некотором внутреннем интервале $a_1 \leqslant x \leqslant b_1$, где $a \leqslant a_1 < b_1 \leqslant b$. Тогда $a_1 \leqslant x \leqslant b_1$ называется *интервалом сходимости* ряда (39).

Заметим в заключение, что, как и для числовых рядов (п. III.6), произвольное изменение конечного числа членов ряда (39) не может изменить характера его сходимости.

9. Свойства функциональных рядов. 1. *Сумма равномерно сходящегося ряда из непрерывных функций также не может иметь разрывов*. Действительно, если

$$f_1(x) + f_2(x) + \ldots + f_n(x) + \ldots = S(x) \quad (a \leqslant x \leqslant b), \quad (41)$$

то

$$S(x) = [f_1(x) + \ldots + f_n(x)] + [f_{n+1}(x) + f_{n+2}(x) + \ldots] =$$
$$= S_n(x) + R_n(x). \quad (42)$$

Если члены ряда — непрерывные функции, то $S_n(x)$, как сумма *конечного* числа непрерывных функций, также непрерывна (п. III.14). С другой стороны, если ряд (41) сходится равномерно, то для достаточно большого n остаток $R_n(x)$ будет как угодно мал *на всем* интервале $a \leqslant x \leqslant b$. Поэтому при малом изменении x как $S_n(x)$, так и $R_n(x)$, а потому и вся сумма (42) изменяются мало, т. е. эта сумма не может иметь разрывов.

Иногда рассматриваются ряды вида (41) на конечном или бесконечном интервале $a < x < b$, которые равномерно сходятся не на

всем этом интервале, а на *каждом «строго внутреннем» интервале* $a_1 \leqslant x \leqslant b_1$, где $a < a_1 < b_1 < b$. Тогда доказанное свойство можно применить к интервалу $a_1 \leqslant x \leqslant b_1$, а затем, приближая a_1 к a, а b_1 к b, заключить, что сумма ряда не имеет разрывов и на всем интервале $a < x < b$. Аналогичное замечание справедливо и для дальнейших свойств.

Если члены ряда (41) имеют разрывы, то, рассуждая подобным образом, получим, что *если ряд* (41) *сходится равномерно, то его сумма может иметь разрывы только в тех точках, в которых имели разрывы слагаемые.* В отличие от этого, если ряд сходится в среднем, его сумма может иметь новые разрывы; она может иметь разрывы, даже если все члены ряда были непрерывными. Это связано с тем, что непрерывные функции $S_n(x)$ могут сходиться в смысле среднего интегрального к разрывной функции (рис. 270).

2. *Ряд, сходящийся равномерно, можно почленно интегрировать*, т. е. из (41) вытекает, что

$$\int_{x_0}^{x} f_1(t)\,dt + \int_{x_0}^{x} f_2(t)\,dt + \ldots + \int_{x_0}^{x} f_n(t)\,dt + \ldots = \int_{x_0}^{x} S(t)\,dt,$$

причем полученный ряд сходится равномерно на интервале $a \leqslant x \leqslant b$. Действительно,

$$\left|\int_{x_0}^{x} S(t)\,dt - \sum_{k=1}^{n}\int_{x_0}^{x} f_k(t)\,dt\right| = \left|\int_{x_0}^{x}\left[S(t) - \sum_{k=1}^{n} f_k(t)\right]dt\right| =$$

$$= \left|\int_{x_0}^{x}[S(t) - S_n(t)]\,dt\right| \leqslant \left|\int_{x_0}^{x}|S(t) - S_n(t)|\,dt\right| \leqslant$$

$$\leqslant \int_{a}^{b}|S(t) - S_n(t)|\,dt \leqslant (b-a)\cdot \max_{a \leqslant t \leqslant b}|S(t) - S_n(t)| \underset{n\to\infty}{\longrightarrow} 0.$$

То же доказательство, с отброшенным последним членом в цепи неравенств, показывает, что и ряд, сходящийся в среднем, можно почленно интегрировать, причем полученный после интегрирования ряд сходится равномерно на интервале $a \leqslant x \leqslant b$

3. *Ряд из непрерывных функций, сходящийся равномерно, можно почленно дифференцировать, если после этого получится ряд, сходящийся равномерно*, т. е. из (41) следует, что

$$f_1'(x) + f_2'(x) + \ldots + f_n'(x) + \ldots = S'(x).$$

Для доказательства надо обозначить сумму последнего ряда через $Q(x)$ и произвести его почленное интегрирование на основе свойства 2; получится равенство

$$S(x) - S(x_0) = \int_{x_0}^{x} Q(t)\,dt,$$

дифференцируя которое, найдем, что $Q(x) = S'(x)$.

Если уточнить понятие сходимости функционального ряда, воспользовавшись обобщенными функциями (п. XIV.27), то все оговорки о характере сходимости отпадают, т. е. любые сходящиеся ряды будет возможно почленно интегрировать и дифференцировать любое число раз.

§ 3. Степенные ряды

10. Интервал сходимости. *Степенной ряд* имеет вид

$$a_0+a_1x+a_2x^2+\ldots+a_nx^n+\ldots \tag{43}$$

Такие ряды уже встречались в нашем курсе, начиная с п. IV.16. Излагая общую теорию этих рядов, мы для простоты предположим, что существует конечный или бесконечный предел

$$\lim_{n\to\infty}\frac{|a_n|}{|a_{n+1}|}=R, \tag{44}$$

хотя окончательные результаты справедливы в самом общем случае.

Легко выяснить, для каких численных значений x ряд (43) сходится. Так как

$$\lim_{n\to\infty}\frac{|a_{n+1}x^{n+1}|}{|a_nx^n|}=\lim_{n\to\infty}\frac{|x|}{|a_n/a_{n+1}|}=\frac{|x|}{|R|}, \tag{45}$$

то по признаку Даламбера (п. 2) при $|x|<R$, т. е. *при*

$$-R<x<R, \tag{46}$$

ряд (43) *абсолютно сходится*; интервал (46) служит *интервалом сходимости* степенного ряда (43), а R называется *радиусом сходимости. При* $|x|>R$, т. е. при $-\infty<x<-R$ или $R<x<\infty$, вне интервала сходимости, предел (45) больше единицы, т. е. *ряд* (43) *расходится.* При $x=\pm R$, т. е. на концах интервала сходимости, предел (45) равен 1 и признак Даламбера неприменим. В разных примерах здесь может получиться сходимость или расходимость.

Если предела (44) не существует, то и тогда интервал сходимости иногда можно выяснить по признаку Даламбера.

Например, для сходимости ряда

$$1-\frac{x^3}{2\cdot 2^2}+\frac{x^6}{3\cdot 2^4}-\frac{x^9}{4\cdot 2^6}+\frac{x^{12}}{5\cdot 2^8}-\ldots,$$

где предела (44) не существует (почему?), должно быть

$$\lim\left\{\frac{|x^{3(n+1)}|}{(n+2)\,2^{2(n+1)}}:\frac{|x^{3n}|}{(n+1)\,2^{2n}}\right\}=\frac{|x|^3}{2^2}<1.$$

Значит, ряд сходится при $|x^3|<2^2=4$, т. е. интервал сходимости его таков: $-\sqrt[3]{4}<x<\sqrt[3]{4}$. На концах этого интервала при $x=-\sqrt[3]{4}$ ряд расходится, а при $x=\sqrt[3]{4}$ неабсолютно сходится (проверьте!).

Если признак Даламбера применить нельзя, то и тогда можно доказать, что ряд (43) сходится на некотором интервале вида (46), хотя R найти сложнее.

Если $R=\infty$, то ряд (43) сходится при всех x, т. е. на всей оси, хотя при больших $|x|$ сходимость будет очень плохой (см. конец п. 4). Теоретически возможен также случай $R=0$, но тогда ряд (43) сходится только для $x=0$, и потому мы такие ряды здесь не будем рассматривать.

Проверьте, что *для разложений* (IV.55—61) *радиусы сходимости соответственно равны* $R=\infty$, ∞, ∞, ∞, ∞, 1 и 1.

Рассматриваются также степенные ряды вида

$$a_0+a_1(x-a)+a_2(x-a)^2+\ldots+a_n(x-a)^n+\ldots \quad (47)$$

Если обозначить $x-a=x_1$, то видно, что ряд сходится при

$$-R<x-a<R, \quad \text{т. е.} \quad a-R<x<a+R.$$

11. Свойства степенных рядов. 1. *Ряд*

$$a_0+a_1x+a_2x^2+\ldots+a_nx^n+\ldots \quad (48)$$

сходится равномерно (п. 8) *на любом интервале* $-R_1\leqslant x\leqslant R_1$, *где* $0<R_1<R$; R — радиус сходимости ряда (48). Действительно, на таком интервале можно написать

$$|a_0|=|a_0|, \quad |a_1x|\leqslant|a_1R_1|, \quad |a_2x^2|\leqslant|a_2R_1^2|, \quad |a_3x^3|\leqslant|a_3R_1^3|, \quad \ldots,$$

т. е. члены ряда (48) по абсолютной величине не превосходяг соответствующих членов числового ряда

$$|a_0|+|a_1R_1|+|a_2R_1^2|+|a_3R_1^3|+\ldots$$

Последний ряд сходится, так как R_1 лежит внутри интервала сходимости. Значит, по признаку Вейерштрасса (п. 8) ряд (48) равномерно сходится на указанном интервале.

Равномерной сходимости сразу на всем интервале $-R<x<R$ в общем случае не будет. Однако Абель доказал, что *если ряд* (48) *сходится на каком-либо из концов интервала сходимости, то интервал, на котором гарантируется равномерная сходимость, можно считать продолженным до этого конца.*

2. *Сумма ряда* (48) *непрерывна внутри его интервала сходимости.* Действительно, это следует из свойства 1 п. 9. Кроме того, из указанной только что теоремы Абеля вытекает, что если ряд (48) сходится на каком-либо из концов интервала сходимости, то сумма будет непрерывной и в этом конце.

3. *При почленном интегрировании или дифференцировании ряда* (48) *его радиус сходимости не меняется.* Например, при

почленном интегрировании получится ряд

$$a_0x+\frac{a_1}{2}x^2+\frac{a_2}{3}x^3+\ldots+\frac{a_{n-1}}{n}x^n+\frac{a_n}{n+1}x^{n+1}+\ldots$$

Вычислим его радиус сходимости по формуле (44):

$$\lim_{n\to\infty}\frac{|a_{n-1}|/n}{|a_n|/(n+1)}=\lim_{n\to\infty}\frac{(n+1)}{n}\frac{|a_{n-1}|}{|a_n|}=\lim_{n\to\infty}\frac{n+1}{n}\lim_{n\to\infty}\frac{|a_{n-1}|}{|a_n|}=1\cdot R,$$

т. е. получается то же значение (44).

4. *Формулу*

$$a_0+a_1x+\ldots+a_nx^n+\ldots=S(x)\qquad(-R<x<R)$$

можно почленно интегрировать и дифференцировать любое число раз на интервале сходимости. Это вытекает из только что доказанных свойств 1 и 3 и из свойств 2 и 3 п. 9, так как если при однократном дифференцировании или интегрировании радиус сходимости не меняется, то он не меняется и при повторении этих действий.

Из свойств 4 и 2 вытекает, в частности, что *сумма степенно́го ряда имеет внутри своего интервала сходимости непрерывные производные всех порядков.*

Рассмотрим, например, ряд

$$\frac{1}{1+x^2}=1-x^2+x^4-x^6+x^8-\ldots\qquad(-1<x<1);$$

его можно получить из ряда (IV.60), положив $a=-1$, либо же просто воспользовавшись формулой (III.7) для суммы бесконечной убывающей геометрической прогрессии. Почленно интегрируя, получим

$$\int_0^x\frac{1}{1+x^2}\,dx=\operatorname{arctg}x=x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+\ldots\qquad(-1<x<1).\qquad(49)$$

Так как ряд в правой части сходится и при $x=1$, то в силу теоремы Абеля формула (49) справедлива и при $x=1$, т. е. мы находим сумму интересного числового ряда

$$1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\ldots=\operatorname{arctg}1=\frac{\pi}{4}.$$

При помощи почленных интегрирований и дифференцирований иногда удается свести заданный ряд к известным рядам и тем самым найти его сумму. Например, найдем сумму ряда

$$2+\frac{3}{1!}x+\frac{4}{2!}x^2+\frac{5}{3!}x^3+\ldots=S(x).$$

На основании признака Даламбера легко проверить, что ряд сходится на всей оси, т. е. $R=\infty$. Умножим обе части на x и проинтегрируем результат от нуля до некоторого x:

$$\int_0^x xS(x)\,dx=x^2+\frac{x^3}{1!}+\frac{x^4}{2!}+\frac{x^5}{3!}+\ldots=x^2\left(1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots\right)=x^2e^x.$$

Отсюда, дифференцируя, получим

$$xS(x)=(x^2e^x)'=2xe^x+x^2e^x,$$

т. е. окончательно

$$S(x)=(2+x)e^x.$$

Приведем пример иного характера. Найдем сумму ряда

$$\frac{x^2}{1\cdot 2}+\frac{x^3}{2\cdot 3}+\frac{x^4}{3\cdot 4}+\frac{x^5}{4\cdot 5}+\ldots=\sigma(x). \qquad (50)$$

Для этого продифференцируем его почленно:

$$\sigma'(x)=\frac{x}{1}+\frac{x^2}{2}+\frac{x^3}{3}+\frac{x^4}{4}+\ldots=-\ln(1-x)$$

(см. формулу (IV.61)). Отсюда

$$\sigma(x)=-\int\ln(1-x)\,dx=$$

$$=-x\ln(1-x)-\int\frac{x}{1-x}\,dx=x+(1-x)\ln(1-x)+C. \qquad (51)$$

Чтобы выяснить, чему равно в данном случае C, положим в формулах (50) и (51) $x=0$; получим, что $0=\sigma(0)=C$. Окончательно,

$$\sigma(x)=x+(1-x)\ln(1-x).$$

В других примерах часто получаются интегралы, не выражающиеся через элементарные функции.

Иногда в результате аналогичных действий мы получаем дифференциальное уравнение для суммы ряда, решив которое, мы и находим сумму. Найдем, например, сумму ряда

$$\frac{x}{1!}+\frac{x^4}{4!}+\frac{x^7}{7!}+\frac{x^{10}}{10!}+\frac{x^{13}}{13!}+\ldots=p(x). \qquad (52)$$

Для этого трижды продифференцируем формулу (52):

$$1+\frac{x^3}{3!}+\frac{x^6}{6!}+\frac{x^9}{9!}+\frac{x^{12}}{12!}+\ldots=p'(x), \qquad (53)$$

$$\frac{x^2}{2!}+\frac{x^5}{5!}+\frac{x^8}{8!}+\frac{x^{11}}{11!}+\ldots=p''(x), \qquad (54)$$

$$\frac{x}{1!}+\frac{x^4}{4!}+\frac{x^7}{7!}+\frac{x^{10}}{10!}+\ldots=p'''(x).$$

Мы видим, что получился исходный ряд, т. е.

$$p'''(x)-p(x)=0.$$

Решая это уравнение по методу п. XV.17, найдем

$$p(x)=C_1e^x+e^{-\frac{x}{2}}\left(C_2\cos\frac{\sqrt{3}x}{2}+C_3\sin\frac{\sqrt{3}x}{2}\right). \qquad (55)$$

Для вычисления C_1, C_2 и C_3 подставляем $x=0$ в формулы (52), (53) и (54), получим, что $p(0)=0$, $p'(0)=1$, $p''(0)=0$; это — начальное условие для $p(x)$. В силу (55) выводим, что

$$C_1+C_2=0, \quad C_1-\frac{1}{2}C_2+\frac{\sqrt{3}}{2}C_3=1, \quad C_1-\frac{1}{2}C_2-\frac{\sqrt{3}}{2}C_3=0,$$

откуда

$$C_1=\frac{1}{3}, \quad C_2=-\frac{1}{3}, \quad C_3=\frac{1}{\sqrt{3}},$$

и окончательно находим сумму ряда (52):

$$\frac{x}{1!}+\frac{x^4}{4!}+\frac{x^7}{7!}+\frac{x^{10}}{10!}+\ldots=\frac{1}{3}e^x+e^{-\frac{x}{2}}\left(-\frac{1}{3}\cos\frac{\sqrt{3}}{2}x+\frac{1}{\sqrt{3}}\sin\frac{\sqrt{3}}{2}x\right).$$

В некоторых случаях аналогичными способами удается «свернуть» сумму заданного числового ряда в интеграл или даже в простую комбинацию математических констант (целых чисел, π, e и т. п.) и функций от них. Проделаем это, например, с суммой

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\ldots=S. \tag{56}$$

Для этого рассмотрим вспомогательный ряд

$$\frac{x}{1^2}+\frac{x^2}{2^2}+\frac{x^3}{3^2}+\ldots=q(x) \qquad (-1\leqslant x\leqslant 1).$$

При помощи дифференцирования найдем (проверьте!)

$$q(x)=-\int_0^x \frac{\ln(1-x)}{x}\,dx.$$

Отсюда, подставляя $x=1$, получим сумму ряда (56)

$$S=-\int_0^1 \frac{\ln(1-x)}{x}\,dx. \tag{57}$$

Соответствующий неопределенный интеграл не является элементарной функцией, однако все же иногда предпочтительнее иметь ответ в «свернутом» виде, а не в виде ряда. Кстати, в п. 25 мы совершенно другим методом найдем, что сумма ряда (56) равна $\frac{\pi^2}{6}$; отсюда, в частности, получается значение интеграла (57).

12. Конечные действия над степенными рядами. Так как степенные ряды внутри своих интервалов сходимости абсолютно сходятся, то их можно, в частности, *почленно складывать*, *множить на общий множитель* (п. 3) *и множить друг на друга* по правилу умножения многочленов (п. 6).

Рассмотрим, например, как происходит умножение рядов:

$$e^x \ln(1+x) = \left(1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots\right)\left(x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\ldots\right)=$$
$$=x+\left(\frac{1}{1!}-\frac{1}{2}\right)x^2+\left(\frac{1}{2!}-\frac{1}{2\cdot 1!}+\frac{1}{3}\right)x^3+$$
$$+\left(\frac{1}{3!}-\frac{1}{2\cdot 2!}+\frac{1}{3\cdot 1!}-\frac{1}{4}\right)x^4+\left(\frac{1}{4!}-\frac{1}{2\cdot 3!}+\frac{1}{3\cdot 2!}-\frac{1}{4\cdot 1!}+\frac{1}{5}\right)x^5+\ldots$$
$$\ldots = x+\frac{1}{2}x^2+\frac{1}{3}x^3+\frac{1}{6}x^4+\frac{3}{40}x^5+\ldots,$$

причем здесь можно подсчитать столько коэффициентов, сколько понадобится. У первого ряда радиус сходимости равен ∞, у второго равен 1, т. е. полученный результат справедлив при $-1 < x < 1$, где абсолютно сходятся оба ряда.

Аналогично осуществляется *деление ряда на ряд*, которое мы покажем на примере

$$\operatorname{tg} x = \frac{\sin x}{\cos x} = \frac{\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\ldots}{1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots} = \frac{\frac{x}{1}-\frac{x^3}{6}+\frac{x^5}{120}-\frac{x^7}{5040}+\ldots}{1-\frac{x^2}{2}+\frac{x^4}{24}-\frac{x^6}{720}+\ldots};$$

$$\begin{array}{l|l}
x-\frac{x^3}{6}+\frac{x^5}{120}-\frac{x^7}{5040}+\ldots & 1-\frac{x^2}{2}+\frac{x^4}{24}+\frac{x^6}{720}-\ldots \\
\underline{-\left(x-\frac{x^3}{2}+\frac{x^5}{24}-\frac{x^7}{720}+\ldots\right)} & \overline{x+\frac{x^3}{3}+\frac{2}{15}x^5+\frac{17}{315}x^7+\ldots} \\
\quad\frac{x^3}{3}-\frac{x^5}{30}+\frac{x^7}{840}+\ldots & \\
\quad\underline{-\left(\frac{x^3}{3}-\frac{x^5}{6}+\frac{x^7}{72}+\ldots\right)} & \\
\qquad\frac{2x^5}{15}-\frac{4x^7}{315}+\ldots & \\
\qquad\underline{-\left(\frac{2x^5}{15}-\frac{x^7}{15}+\ldots\right)} & \\
\qquad\quad\frac{17}{315}x^7+\ldots & \\
\qquad\quad\underline{-\left(\frac{17}{315}x^7+\ldots\right)} & \\
\qquad\quad\ldots\ldots\ldots\ldots &
\end{array}$$

Итак, разложение тангенса в степенной ряд начинается с членов

$$\operatorname{tg} x = x+\frac{x^3}{3}+\frac{2}{15}x^5+\frac{17}{315}x^7+\ldots \qquad (58)$$

Для вычисления дальнейших членов надо было продолжить разложения $\sin x$ и $\cos x$. Можно доказать, что формула (58) справедлива при $|x| < \frac{\pi}{2}$.

Разложение (58) можно получить также с помощью метода неопределенных коэффициентов. Для этого заметим, что $\operatorname{tg} x$ как нечетная функция

должна разлагаться в ряд по нечетным степеням

$$\operatorname{tg} x = a_1 + a_3x^3 + a_5x^5 + a_7x^7 + \ldots$$

Но так как $\cos x \cdot \operatorname{tg} x = \sin x$, то

$$\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots\right)(a_1x + a_3x^3 + a_5x^5 + a_7x^7 + \ldots) =$$
$$= \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

Раскрывая скобки и приравнивая коэффициенты при одинаковых степенях x, получим

$$a_1 = \frac{1}{1!}; \quad a_3 - \frac{a_1}{2!} = -\frac{1}{3!}; \quad a_5 - \frac{a_3}{2!} + \frac{a_1}{4!} = \frac{1}{5!}; \quad a_7 - \frac{a_5}{2!} + \frac{a_3}{4!} - \frac{a_1}{6!} = -\frac{1}{7!}, \ldots,$$

откуда последовательно найдем коэффициенты a_1, a_3, a_5, ...

Наконец, применяется *подстановка ряда*

$$y = f(x) = a_0 + a_1x + a_2x^2 + \ldots$$

в ряд

$$\varphi(y) = b_0 + b_1y + b_2y^2 + \ldots \tag{59}$$

или, в более общем случае, в ряд

$$\psi(y) = c_0 + c_1(y - a) + c_2(y - a)^2 + \ldots$$

Например, для ряда (59) это даст

$$\varphi(f(x)) = b_0 + b_1(a_0 + a_1x + a_2x^2 + \ldots) + b_2(a_0 + a_1x + a_2x^2 + \ldots)^2 + \ldots,$$

причем в правой части надо раскрыть скобки и привести подобные члены. Чтобы результат имел смысл при $x = 0$, нужно, чтобы ряд (59) сходился при $y = a_0$ (почему?), т. е. чтобы a_0 попало на интервал сходимости ряда (59); впрочем, если это условие не выполнено, то вычисления сами покажут ошибку.

Приведем пример:

$$\ln(1 + \sin x) = \frac{\sin x}{1} - \frac{(\sin x)^2}{2} + \frac{(\sin x)^3}{3} - \ldots =$$
$$= \frac{x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \ldots}{1} - \frac{\left(x - \frac{x^3}{6} + \frac{x^5}{120} - \ldots\right)^2}{2} +$$
$$+ \frac{\left(x - \frac{x^3}{6} + \frac{x^5}{120} - \ldots\right)^3}{3} - \ldots = \frac{x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \ldots}{1} -$$
$$- \frac{x^2 - \frac{x^4}{3} + \frac{2}{45}x^6 + \ldots}{2} + \frac{x^3 - \frac{x^5}{2} + \frac{13}{120}x^7 + \ldots}{3} -$$
$$- \frac{x^4 - \frac{2}{3}x^6 + \ldots}{4} + \frac{x^5 - \frac{5}{6}x^7 + \ldots}{5} - \frac{x^6 + \ldots}{6} + \frac{x^7 + \ldots}{7} + \ldots;$$

здесь при вычислении последовательных степеней ряда мы множили очередную степень на основной ряд для $\sin x$ по правилу умножения многочленов,

причем слишком высокие степени x — в данном примере начиная с x^8 — просто отбрасывали. Приводя подобные члены, получим окончательно

$$\ln(1+\sin x)=x-\frac{x^2}{2}+\frac{x^3}{6}-\frac{x^4}{12}+\frac{x^5}{24}-\frac{x^6}{45}+\frac{61x^7}{5040}+\ldots$$

При помощи методов пп. 11 — 12 оказывается возможным, исходя из простейших рядов (п. IV.16), получить разложения многих других функций. Часто оказывается затруднительным написать выражение для общего члена, однако всегда можно вычислить любое число первых членов, чего обычно бывает достаточно.

13. Степенной ряд как ряд Тейлора. Рассмотрим сумму ряда

$$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n+\ldots \quad (-R<x<R). \tag{60}$$

Коэффициенты этого ряда легко выразить через его сумму. Для этого, как в п. IV.15, будем последовательно дифференцировать формулу (60) и подставлять $x=0$; получится

$$\begin{aligned} & & f(0)&=a_0;\\ f'(x)&=1a_1+2a_2x+3a_3x^2+\ldots, & f'(0)&=1a_1;\\ f''(x)&=1\cdot 2a_2+2\cdot 3a_3x+3\cdot 4a_4x^2+\ldots, & f''(0)&=1\cdot 2a_2;\\ f'''(x)&=1\cdot 2\cdot 3a_3+2\cdot 3\cdot 4a_4x+3\cdot 4\cdot 5a_5x^2+\ldots, & f'''(0)&=1\cdot 2\cdot 3a_3 \end{aligned}$$

и т. д.

Находя отсюда a_0, a_1, a_2, ... и подставляя их в (60), получим

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\ldots$$
$$\ldots+\frac{f^{(n)}(0)}{n!}x^n+\ldots \quad (-R<x<R), \tag{61}$$

т. е. уже знакомый нам ряд Тейлора (IV.54). Итак, *степенной ряд есть ряд Тейлора для своей суммы.*

Так как коэффициенты ряда (61) однозначно выражаются через его сумму, то получаем в частности, что *если суммы двух степенных рядов тождественно совпадают, то их коэффициенты при одинаковых степенях x также равны; если сумма степенного ряда тождественно равна нулю, то и все его коэффициенты равны нулю*

Выше мы считали, что степенной ряд задан. На практике часто ставится противоположная задача о разложении заданной функции $f(x)$ в степенной ряд. При этом возникает вопрос, на каком интервале будет справедлива формула (61). На основании пп. IV.15 — 16 это равносильно вопросу о том, при каких x остаточный член конечной формулы Тейлора стремится к нулю при возрастании номера.

Непосредственно исследовать этот остаточный член удается лишь в редких случаях. К счастью, такое исследование является излишним: можно доказать, что, например, для элементарной функции $f(x)$ формула (61) имеет место на всяком интервале; на котором ряд сходится если все значения $f(0)$, $f'(0)$ $f'''(0)$ ... получаются

непосредственной подстановкой значения $x=0$ в выражения для $f(x)$, $f'(x)$, $f''(x)$, ... Отсюда, в частности, вытекает, что разложения, приведенные в п. IV.16, имеют место на интервалах сходимости соответствующих рядов.

В то же время далеко не каждую функцию можно разложить в степенной ряд (ряд Тейлора). Например, степенной ряд не может представлять функцию на интервале, на котором она или производная от нее какого-либо порядка имеют разрывы, а также если она на различных частях этого интервала задается различными формулами (п. I.13).

Все сказанное непосредственно переносится на ряды по степеням $x-a$ вида (47) и на соответствующие ряды Тейлора (IV.53).

14. Степенные ряды с комплексными членами. Эти ряды имеют вид

$$a_0+a_1z+a_2z^2+\ldots+a_nz^n+\ldots,\quad z=x+iy, \tag{62}$$

где коэффициенты a_n и независимая переменная z принимают любые комплексные значения. Теория таких рядов совершенно аналогична теории вещественных степенных рядов, однако неравенство $|z|<R$, выделяющее те z, для которых ряд (62) сходится, определяет на плоскости z *круг сходимости* ряда (62) (рис. 338). Подобным образом, для ряда по степеням $z-a$, где a — какое-нибудь комплексное число, неравенство $|z-a|<R$ определяет круг радиуса R с центром в точке a. Если $R=\infty$, то ряд сходится на всей комплексной плоскости. Свойства, указаные в п. 11 и 12, без существенных изменений переносятся на ряды вида (62); при этом сумма $S(z)$ будет *комплексной функцией от комплексного переменного* (п. VIII.11); интеграл понимается как неопределенный (первообразная). Примеры определения функций для комплексных значений аргумента с помощью рядов вида (62) были приведены в п. VIII.4.

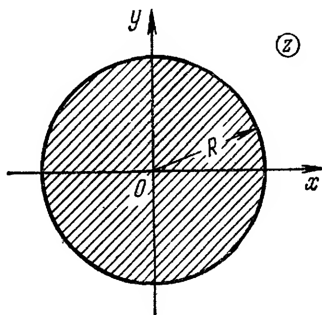


Рис. 338.

В п. VIII.4 мы говорили, что при этом тождества, справедливые для вещественных значений аргумента, сохраняют силу и для комплексных его значений. Покажем это на примере равенства

$$e^{\ln(1+x)}=1+x. \tag{63}$$

Для вещественных x оно справедливо в силу определения логарифма. Значит, если в ряд для e^y подставить ряд для $y=\ln(1+x)$ (п. 12), то после тождественных преобразований получится $1+x$. Если провести эти же преобразования, но считать x комплексным, т. е. писать z вместо x, то получится

$$e^{\ln(1+z)}=1+z, \tag{64}$$

что и требовалось доказать.

Из этой формулы, в частности, следует, что определение логарифма с помощью степенного ряда (IV.61), где вместо x надо писать z, соответствует определению логарифма комплексного числа, данному в п. VIII.5. При этом сумма ряда дает лишь одну ветвь бесконечнозначной логарифмической функции, именно ту, которая при $z=1$ обращается в нуль.

Подобным образом доказываются формула (VIII.7) и другие аналогичные формулы.

15. Понятие о числах Бернулли. В теории рядов, в частности степенных, а также в других отделах математики применяются так называемые *числа Бернулли*, открытые Я. Бернулли. Эти числа β_1, β_2, β_3, β_4, ... определяются с помощью символического рекуррентного соотношения

$$(\beta+1)^{n+1}-\beta^{n+1}=0 \quad (n=1,\ 2,\ 3,\ \ldots),$$

где в левой части после раскрытия скобок надо в каждом члене заменить β^k на β_k.

Покажем, как это делается, на первых номерах:

$$n=1;\quad \beta_2+2\beta_1+1-\beta_2=0,\ \text{т. е.}\ \beta_1=-\frac{1}{2};$$

$$n=2;\quad \beta_3+3\beta_2+3\beta_1+1-\beta_3=0,\ \text{т. е.}\ \beta_2=-\frac{3\beta_1+1}{3}=\frac{1}{6};$$

$$n=3;\quad \beta_4+4\beta_3+6\beta_2+4\beta_1+1-\beta_4=0,\ \text{т. е.}\ \beta_3=-\frac{6\beta_2+4\beta_1+1}{4}=0.$$

Дальнейшие вычисления дают

$$\beta_4=-\frac{1}{30},\quad \beta_5=0,\quad \beta_6=\frac{1}{42},\quad \beta_7=0,\quad \beta_8=-\frac{1}{30},\ \beta_9=0,$$

$$\beta_{10}=\frac{5}{66},\quad \beta_{11}=0,\quad \beta_{12}=-\frac{691}{2730},\quad \beta_{13}=0,\quad \beta_{14}=\frac{7}{6},\ \ldots$$

Можно доказать, что все β_n с нечетными $n\geqslant 3$ равны нулю. Обозначив

$$B_n=(-1)^{n-1}\beta_{2n} \qquad (n=1,\ 2,\ 3,\ \ldots),$$

получим

$$B_1=\frac{1}{6},\quad B_2=\frac{1}{30},\quad B_3=\frac{1}{42},\quad B_4=\frac{1}{30},\quad B_5=\frac{5}{66},\quad B_6=\frac{691}{2730},\ \ldots$$

Эти числа также называются числами Бернулли.

Приведем некоторые формулы, содержащие числа Бернулли:

$$\sum_{n=1}^{\infty}\frac{1}{n^{2k}}=\zeta(2k)\ \ (\text{см. п. } 4)=\frac{B_k(2\pi)^{2k}}{2\,(2k)!} \qquad (k=1,\ 2,\ \ldots); \tag{65}$$

в частности,

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{B_1(2\pi)^2}{2\cdot 2!}=\frac{\pi^2}{6},\qquad \sum_{n=1}^{\infty}\frac{1}{n^4}=\frac{B_2(2\pi)^4}{2\cdot 4!}=\frac{\pi^4}{90}\ \text{ и т. д.}$$

Из формулы (65) вытекает, что все числа B_n положительные. Далее,

$$\operatorname{tg} x=\sum_{n=1}^{\infty}\frac{2^{2n}(2^{2n}-1)}{(2n)!}B_n x^{2n-1}$$

(ср. с формулой (58)) и т. д.

16. Применение рядов к решению разностных уравнений. *Разностное уравнение* связывает неизвестную величину и ее разности (п. V.7). Рассмотрим сначала случай, когда искомой является последовательность a_0, a_1, a_2,... ..., a_n, ... Тогда разностное уравнение, скажем, *второго порядка* в общем случае имеет вид

$$f(n,\ a_n,\ \Delta a_n,\ \Delta^2 a_n)=0 \qquad (n=0,\ 1,\ 2,\ \ldots), \tag{66}$$

где $\Delta a_n = a_{n+1}-a_n$, $\Delta^2 a_n = \Delta a_{n+1}-\Delta a_n$. Если подставить

$$\Delta a_n = a_{n+1}-a_n, \quad \Delta^2 a_n = a_{n+2}-2a_{n+1}+a_n,$$

то взамен (66) получится уравнение вида

$$\varphi(n,\ a_n,\ a_{n+1},\ a_{n+2})=0 \quad (n=0,\ 1,\ 2,\ \ldots). \tag{67}$$

Конечно, в частных случаях в левые части уравнений (66) и (67) могут входить не все выписанные там аргументы.

Чтобы решить уравнение (67), можно, например, произвольно задать a_0 и a_1; затем, положив $n=0$ в (67), найти a_2; далее, положив $n=1$ в (67) и подставив найденное значение a_2, найти a_3 и т. д Таким *методом шагов* можно найти сколько угодно членов последовательности a_n.

Если уравнение (67) является линейным однородным с постоянными коэффициентами, т е. имеет вид

$$\alpha a_n+\beta a_{n+1}+\gamma a_{n+2}=0 \qquad (n=0,\ 1,\ 2,\ \ldots;\ \alpha,\ \beta,\ \gamma=\text{const}), \tag{68}$$

то решение можно найти в общем виде, как мы сейчас покажем; аналогичный метод применим для уравнений любого порядка.

Рассмотрим *производящий* степенной ряд для искомой последовательности, т. е. ряд

$$Q=a_0+a_1x+a_2x^2+\ldots+a_nx^n+\ldots,$$

коэффициенты которого равны ее членам. Мы будем интересоваться только коэффициентами этого ряда и не будем придавать x никаких числовых значений; тогда степенной ряд называется *формальным*. Легко найти произведение

$$(\gamma+\beta x+\alpha x^2)\,Q=$$
$$=\gamma a_0+(\beta a_0+\gamma a_1)\,x+(\alpha a_0+\beta a_1+\gamma a_2)\,x^2+(\alpha a_1+\beta a_2+\gamma a_3)\,x^3+\ldots$$

В правой части все коэффициенты, начиная с коэффициента при x^2, равны нулю в силу уравнения (68). Производя деление, получим

$$Q=\frac{\gamma a_0+(\beta a_0+\gamma a_1)\,x}{\gamma+\beta x+\alpha x^2}. \tag{69}$$

Так как a_0 и a_1 заданы, то в правой части получается отношение двух многочленов с заданными коэффициентами. Его можно разложить на сумму элементарных дробей первого типа по методу п.VIII.10. Каждая из них имеет вид

$$\frac{A}{(x-a)^\alpha}=\frac{A}{(-a)^\alpha\left(1-\dfrac{x}{a}\right)^\alpha}=\frac{B}{(1-\gamma x)^\alpha} \qquad \left(B=\frac{A}{(-a)^\alpha};\quad \gamma=\frac{1}{a}\right);$$

для дроби (69) $\alpha=1$ или 2, но для разностных уравнений высшего порядка α может получиться бо́льшим. Эти дроби разлагаем в степенные ряды по формулам, получающимся из суммы геометрической прогрессии при помощи

дифференцирования:

$$\frac{B}{1-\gamma x}=B+B\gamma x+B\gamma^2x^2+\ldots+B\gamma^n x^n+\ldots,$$

$$\frac{B}{(1-\gamma x)^2}=\frac{1}{\gamma}\left(\frac{B}{1-\gamma x}\right)'=B+2B\gamma x+3B\gamma^2x^2+\ldots+(n+1)\ B\gamma^n x^n+\ldots$$

и т. д. Суммируя коэффициенты при x^n у всех полученных рядов, мы тем самым находим коэффициент a_n ряда Q, т. е. решение уравнения (68) в общем виде.

В качестве примера мы предлагаем читателю вывести общую формулу для *чисел Фибоначчи* $a_0=0$, $a_1=1$, $a_2=1$, $a_3=2$, $a_4=3$, $a_5=5$, ..., каждое из которых, начиная с третьего, равно сумме двух предшествующих:

$$a_n=\frac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{\sqrt{5}\cdot 2^n}.$$

Рассматриваются также разностные уравнения, в которых искомой является функция $y(x)$, т. е. когда взамен (66) и (67) уравнение имеет одну из форм

$$f(x,\ y,\ \Delta_h y,\ \Delta_h^2 y)=0,$$

$$\varphi(x,\ y(x),\ y(x+h),\ y(x+2h))=0. \tag{70}$$

Этот случай сводится к предшествующему, в котором искомой является последовательность. Пусть, например, $0\leqslant x<\infty$. Тогда надо обозначить $a_n=$ $=y(\xi+nh)$ $(n=0,\ 1,\ 2,\ \ldots)$, где ξ — какое-либо постоянное число, $0\leqslant\xi<h$. Если в уравнении (70) положить $x=\xi+nh$, то оно перепишется в виде

$$\varphi(\xi+nh,\ a_n,\ a_{n+1},\ a_{n+2})=0,$$

т. е. при постоянном ξ будет иметь форму (67). Найдя a_n и пользуясь произволом в выборе ξ, мы и получаем искомое решение $y(x)$. В частности, отсюда следует, что для уравнения (70) значения $y(x)$ при $0\leqslant x<2h$ можно задавать произвольно (почему?).

17. Кратные степенные ряды. Кратные степенные ряды играют для функций нескольких переменных такую же роль, как простые степенные ряды для функций одной переменной. Рассмотрим для простоты двойные степенные ряды; рассмотрение степенных рядов высшей кратности проходит аналогично.

При записи двойного степенного ряда, как и при записи двойного числового ряда (п. 6), удобно пользоваться двойными индексами:

$$S(x,\ y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}a_{mn}x^my^n=a_{00}+a_{10}x+a_{01}y+a_{20}x^2+a_{11}xy+a_{02}y^2+\ldots \tag{71}$$

От такого ряда требуется абсолютная сходимость, так что порядок его суммирования несуществен.

Область сходимости ряда (71) — это некоторая область на плоскости x, y (ею может оказаться, в частности, вся эта плоскость); эта область может иметь вид, изображенный на рис. 339. При каждом зафиксированном y получается ряд по степеням x, радиус сходимости которого R может зависеть от y, т. е. $R=R(y)\leqslant\infty$. Значит, область сходимости симметрична относительно оси y; аналогично проверяем симметрию относительно оси x. Так как рассматривается абсолютная сходимость ряда (71), то $R(y)$ при $y\geqslant 0$ является невозрастающей функцией y (почему?).

Аналогично рассматриваются ряды вида

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{mn}(x-a)^m (y-b)^n. \qquad (72)$$

Здесь получается область сходимости с центром симметрии в точке $(a;\ b)$

Свойства кратных степенных рядов подобны свойствам простых степенных рядов (п.11—12). Кратные ряды получаются, в частности, при разложении функции нескольких переменных в ряд Тейлора (п. XII.6).

$$f(x,\ y)=f(0,\ 0)+\frac{f'_x(0,\ 0)}{1!}x+\frac{f'_y(0,\ 0)}{1!}y+\frac{f''_{xx}(0,\ 0)}{2!}x^2+\frac{f''_{xy}(0,\ 0)}{1!1!}xy+\frac{f''_{yy}(0,\ 0)}{2!}y^2+\ldots;$$

аналогично получаются ряды (72). Они рассматриваются также при применении метода малого параметра (пп. V.5 и XV.27), если уравнение содержит несколько параметров, и в других вопросах.

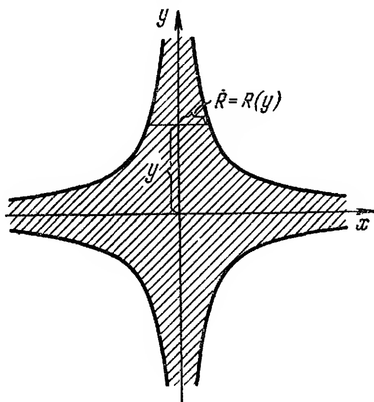


Рис. 339.

18. Функции от матриц. Пусть **A** — квадратная матрица (гл. XI), для определенности, третьего порядка; результаты, которые мы получим, справедливы для матриц любого порядка. В пп. XI.2—3 мы видели, что такое $\mathbf{A}^2$ и $\mathbf{A}^{-1}$. Но что́ надо понимать, например, под $e^{\mathbf{A}}$? Значение экспоненциальной функции в математике подсказывает, что этот вопрос далеко не праздный.

Оказывается, что разумный ответ на поставленный вопрос можно дать с помощью степенных рядов, наподобие того как в п. VIII.4 были определены функции от комплексного переменного. Пусть некоторая функция $f(x)$ допускает разложение в ряд

$$f(x)=a_0+a_1x+a_2x^2+a_3x^3+\ldots+a_nx^n+\ldots \qquad (73)$$

Тогда *по определению*

$$f(\mathbf{A})=a_0\mathbf{I}+a_1\mathbf{A}+a_2\mathbf{A}^2+\ldots+a_n\mathbf{A}^n+\ldots, \qquad (74)$$

где **I** — единичная матрица того же порядка, что и **A**. Например,

$$e^{\mathbf{A}}=\mathbf{I}+\frac{\mathbf{A}}{1!}+\frac{\mathbf{A}^2}{2!}+\frac{\mathbf{A}^3}{3!}+\ldots+\frac{\mathbf{A}^n}{n!}+\ldots \qquad (75)$$

Приведенное определение имеет смысл, если ряд (74) сходится. Можно указать простое условие для этой сходимости. Для этого допустим, что ряд (73) имеет радиус сходимости R, и предположим для простоты, что все собственные значения (п. XI.4) $\lambda_1,\ \lambda_2,\ \lambda_3$ матрицы **A** различны. Тогда, как было показано в п. XI.8, матрицу **A** можно привести к диагональному виду, т. е. существует невырожденная матрица **H**, для которой

$$\mathbf{H}^{-1}\mathbf{A}\mathbf{H}=\Lambda=\operatorname{diag}(\lambda_1,\ \lambda_2,\ \lambda_3).$$

Но тогда

$$\mathbf{A}=\mathbf{H}\Lambda\mathbf{H}^{-1},\quad \mathbf{A}^2=(\mathbf{H}\Lambda\mathbf{H}^{-1})\cdot(\mathbf{H}\Lambda\mathbf{H}^{-1})=\mathbf{H}\Lambda^2\mathbf{H}^{-1},$$
$$\mathbf{A}^3=\mathbf{A}^2\mathbf{A}=(\mathbf{H}\Lambda^2\mathbf{H}^{-1})(\mathbf{H}\Lambda\mathbf{H}^{-1})=\mathbf{H}\Lambda^3\mathbf{H}^{-1} \text{ и т. д.}$$

и ряд (74) можно переписать в виде

$$\mathbf{H}a_0\mathbf{I}\mathbf{H}^{-1}+\mathbf{H}a_1\Lambda\mathbf{H}^{-1}+\mathbf{H}a_2\Lambda^2\mathbf{H}^{-1}+\ldots=$$
$$=\mathbf{H}(a_0\mathbf{I}+a_1\Lambda+a_2\Lambda^2+\ldots)\mathbf{H}^{-1}. \quad (76)$$

Однако диагональную матрицу очень просто возводить в степень:

$$\Lambda^2=\operatorname{diag}(\lambda_1^2,\ \lambda_2^2,\ \lambda_3^2),\qquad \Lambda^3=\operatorname{diag}(\lambda_1^3,\ \lambda_2^3,\ \lambda_3^3) \text{ и т. д.} \quad (77)$$

(проверьте, что вообще при перемножении диагональных матриц получается диагональная матрица, элементы которой равны произведению соответствующих элементов матриц-множителей). Поэтому

$$a_0\mathbf{I}+a_1\Lambda+a_2\Lambda^2+\ldots=$$
$$=\operatorname{diag}(a_0+a_1\lambda_1+a_2\lambda_2^2+\ldots,\ a_0+a_1\lambda_2+a_2\lambda_2^2+\ldots,\ a_0+a_1\lambda_3+a_2\lambda_3^2+\ldots). \quad (78)$$

Если ряды, стоящие на диагонали, сходятся, то сходится и ряд (76), а с ним и ряд (74). Отсюда мы приходим к выводу: *если все собственные значения матрицы* **A** *по модулю меньше* R (радиуса сходимости ряда (73)), *то ряд* (74) *сходится* и притом абсолютно; *если хотя бы одно из этих собственных значений по модулю больше* R, *то ряд* (74) *расходится.* Можно доказать, что этот результат верен и в том случае, когда матрица **A** имеет кратные собственные значения.

Из (76) и (78) вытекает также формула, справедливая для матрицы **A**, приводимой к диагональному виду с помощью матрицы **H**:

$$f(\mathbf{A})=\mathbf{H}\operatorname{diag}(f(\lambda_1),\ f(\lambda_2),\ f(\lambda_3))\mathbf{H}^{-1}.$$

Из доказанного вытекает, например, что ряд (75) сходится для любой матрицы **A**, так как соответствующий ряд (IV.55) имеет бесконечный радиус сходимости. Другой важный ряд

$$\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\ldots+\mathbf{A}^n+\ldots \quad (79)$$

сходится, если все собственные значения матрицы **A** по модулю меньше единицы (почему?).

На функции от матриц распространяются многие свойства обычных функций; эти свойства можно доказывать с помощью действий над рядами, наподобие того как мы вывели формулу (64) из формулы (63). Например, из тождества

$$(1+x+x^2+\ldots)(1-x)=\frac{1}{1-x}(1-x)=1$$

вытекает, что

$$(\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\ldots)(\mathbf{I}-\mathbf{A})=\mathbf{I},$$

т. е. сумма ряда (79) в случае его сходимости равна $(\mathbf{I}-\mathbf{A})^{-1}$. В то же время надо иметь в виду, что при доказательстве ряда свойств с помощью рядов применяется перестановка сомножителей, скажем, равенства вида $ab+ba=2ab$, которая для матриц не всегда возможна. Например, это делается при доказательстве формулы

$$e^{\mathbf{A}}\,e^{\mathbf{B}}=e^{\mathbf{A}+\mathbf{B}},$$

которая, таким образом, *справедлива для перестановочных матриц* **A** *и* **B**.

В качестве примера применения введенных понятий найдем условие сходимости метода последовательных приближений при решении системы линей-

ных алгебраических уравнений. Систему (VI.19) можно в векторно-матричной форме переписать в виде

$$\mathbf{x}=\mathbf{A}\mathbf{x}+\boldsymbol{\delta}, \tag{80}$$

где $\boldsymbol{\delta}$ — заданный вектор, $\mathbf{A}$ — матрица коэффициентов, а $\mathbf{x}$ — искомый вектор. Метод итераций дает, начиная с некоторого $\mathbf{x}=\mathbf{x}_0$:

$$\mathbf{x}_1=\boldsymbol{\delta}+\mathbf{A}\mathbf{x}_0;$$
$$\mathbf{x}_2=\boldsymbol{\delta}+\mathbf{A}\mathbf{x}_1=\boldsymbol{\delta}+\mathbf{A}(\boldsymbol{\delta}+\mathbf{A}\mathbf{x}_0)=\boldsymbol{\delta}+\mathbf{A}\boldsymbol{\delta}+\mathbf{A}^2\mathbf{x}_0;$$
$$\mathbf{x}_3=\boldsymbol{\delta}+\mathbf{A}\mathbf{x}_2=\boldsymbol{\delta}+\mathbf{A}\boldsymbol{\delta}+\mathbf{A}^2\boldsymbol{\delta}+\mathbf{A}^3\mathbf{x}_0$$

и т. д., вообще

$$\mathbf{x}_n=(\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\ldots+\mathbf{A}^{n-1})\,\boldsymbol{\delta}+\mathbf{A}^n\mathbf{x}_0. \tag{81}$$

Для сходимости процесса требуется, чтобы влияние нулевого приближения в пределе сошло на нет, т. е. чтобы $\mathbf{A}^n \underset{n\to\infty}{\longrightarrow} 0$. А для этого в силу формул (77) требуется, чтобы *все собственные значения матрицы* $\mathbf{A}$ *были по модулю меньше единицы*. Это и есть условие сходимости метода итераций. Если оно выполнено, то, переходя в формуле (81) к пределу при $n\longrightarrow\infty$, получим

$$\bar{\mathbf{x}}=\lim_{n\to\infty}\mathbf{x}_n=(\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\ldots+\mathbf{A}^n+\ldots)\,\boldsymbol{\delta}=(\mathbf{I}-\mathbf{A})^{-1}\,\boldsymbol{\delta}.$$

Легко проверить с помощью непосредственной подстановки, что полученный вектор $\bar{\mathbf{x}}$ удовлетворяет уравнению (80) (проделайте это!).

Подобно векторным функциям от скалярного аргумента (п. VII.23) можно рассматривать *матричные функции от скалярного аргумента*, $\mathbf{B}=\mathbf{B}(x)$, на которые легко распространяются многие обычные свойства. Например, часто применяется функция

$$\mathbf{B}=e^{\mathbf{A}x} \qquad (-\infty<x<\infty;\ \mathbf{A}x=x\mathbf{A}),$$

где $\mathbf{A}$ — некоторая постоянная матрица. С помощью рядов легко доказать свойство $(e^{\mathbf{A}x})'=\mathbf{A}e^{\mathbf{A}x}$, из которого, в частности, следует, что для любого постоянного вектора $\mathbf{c}$ будет $(e^{\mathbf{A}x}\mathbf{c})'=\mathbf{A}e^{\mathbf{A}x}\mathbf{c}$. Но это значит, что векторная функция от x,

$$\mathbf{y}=e^{\mathbf{A}x}\mathbf{c}, \tag{82}$$

является решением матричного уравнения (XV.149) с постоянными коэффициентами

$$\mathbf{y}'=\mathbf{A}\mathbf{y}. \tag{83}$$

Если дано начальное условие $\mathbf{y}|_{x=x_0}=\mathbf{y}_0$, то из (82) получаем

$$\mathbf{y}_0=e^{\mathbf{A}x_0}\mathbf{c}, \quad \text{т. е.} \quad \mathbf{c}=e^{-\mathbf{A}x_0}\mathbf{y}_0,$$

и мы получаем *явную формулу для решения*

$$\mathbf{y}=e^{\mathbf{A}x}\,e^{-\mathbf{A}x_0}\,\mathbf{y}_0=e^{(x-x_0)\,\mathbf{A}}\,\mathbf{y}_0.$$

Так как мы смогли удовлетворить любому начальному условию, то формула (82) дает *общее решение уравнения* (83).

19. Асимптотические разложения. Асимптотические разложения, изученные крупным французским математиком А. Пуанкаре́ (1854—1912), широко применяются в настоящее время. Мы будем рассматривать более распространенные ряды по степеням $\frac{1}{x}$, а не x; впрочем, это не принципиально,

так как замена $\frac{1}{x}=x_1$ переводит *разложение на бесконечности* в *разложение в нуле*. Например, из ряда (IV.55) сразу получаем

$$e^{\frac{1}{x}}=1+\frac{1}{1!\,x}+\frac{1}{2!\,x^2}+\ldots+\frac{1}{n!\,x^n}+\ldots \tag{84}$$

Начнем с примера. Рассмотрим функцию

$$f(x)=\int_x^\infty \frac{e^{x-s}}{s}\,ds \qquad (0<x<\infty)$$

при $x\to\infty$. С помощью правила Лопиталя легко проверить, что $f(x)\sim x^{-1}$, т. е. (см. пп. III.8 и III.11)

$$f(x)=\frac{1}{s}+o\left(\frac{1}{x}\right) \qquad (x\to\infty). \tag{85}$$

Чтобы уточнить это разложение, произведем интегрирование по частям:

$$\int_x^\infty \frac{e^{x-s}}{s}\,ds=\left(-e^{x-s}\frac{1}{s}\right)\Bigg|_x^\infty-\int_x^\infty \frac{e^{x-s}}{s^2}\,ds=\frac{1}{x}-\int_x^\infty \frac{e^{x-s}}{s^2}\,ds.$$

Аналогично проверяем, что последний интеграл эквивалентен x^{-2}, т. е получаем

$$f(x)=\frac{1}{x}-\frac{1}{x^2}+o\left(\frac{1}{x^2}\right). \tag{86}$$

Это более точное асимптотическое выражение, чем (85), так как при $x\to\infty$ неопределенное слагаемое в (86), т. е. $o\left(\frac{1}{x^2}\right)$, быстрее стремится к нулю, чем неопределенное слагаемое в (85). Дальнейшие интегрирования по частям дают все более точные асимптотические выражения (проверьте!):

$$f(x)=\frac{1}{x}-\frac{1}{x^2}+\frac{2!}{x^3}+o\left(\frac{1}{x^3}\right); \tag{87}$$

$$f(x)=\frac{1}{x}-\frac{1}{x^2}+\frac{2!}{x^3}-\frac{3!}{x^4}+o\left(\frac{1}{x^4}\right) \text{ и т. д.} \tag{88}$$

Казалось бы, мы должны получить в пределе разложение функции $f(x)$ в ряд

$$\frac{1}{x}-\frac{1!}{x^2}+\frac{2!}{x^3}-\frac{3!}{x^4}+\frac{4!}{x^5}-\ldots \tag{89}$$

Однако с помощью признака Даламбера легко проверить, что этот ряд имеет нулевой радиус сходимости, т. е. *расходится при всех* x! Поэтому рядом (89) «целиком» пользоваться нельзя; но формулы (85) — (88) показывают, что можно пользоваться его частными суммами. Такие ряды называются асимптотически сходящимися.

Общее определение таково. Говорят, что функция $f(x)$ *допускает при* $x\to\infty$ *асимптотическое разложение*

$$f(x)\sim a_0+\frac{a_1}{x}+\frac{a_2}{x^2}+\ldots+\frac{a_n}{x^n}+\ldots, \tag{90}$$

если для любого $n=0, 1, 2, \ldots$ имеет место представление

$$f(x)=a_0+\frac{a_1}{x}+\ldots+\frac{a_n}{x^n}+o\left(\frac{1}{x^n}\right) \qquad (\text{при } x \to \infty).$$

Для обычных разложений в степенные ряды вида (84) это свойство удовлетворяется автоматически. Однако оно может выполняться и в том случае, если ряд (90) всюду расходится (тогда он называется *асимптотически сходящимся* или просто *асимптотическим рядом*) или же сходится, но не к $f(x)$.

При практическом применении ряда (90) его обрывают на некотором члене, а затем, оценивая величину последних членов, заключают, начиная с какого x можно пользоваться выбранной частной суммой.

§ 4. Тригонометрические ряды

20. Свойство ортогональности. Две вещественные функции $g(x)$ и $h(x)$, заданные на конечном или бесконечном интервале $a<x<b$, называются *ортогональными* друг другу на этом интервале, если

$$\int_a^b g(x)\,h(x)\,dx=0. \tag{91}$$

При этом функции предполагаются конечными либо бесконечными, но обязательно с абсолютно сходящимся интегралом (91). Название «ортогональные» объясняется тем, что, как оказывается, формула (91) во многом аналогична условию перпендикулярности двух векторов, заданных своими декартовыми проекциями (п.VII.10).

Система функций называется *ортогональной* на некотором интервале, если каждые две функции из этой системы ортогональны друг другу на этом интервале. Исторически первым и наиболее важным примером ортогональной системы функций явилась система функций

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots,\ \cos nx,\ \sin nx,\ \ldots \tag{92}$$

на интервале $-\pi \leqslant x \leqslant \pi$. Легко убедиться в том, что она действительно ортогональная:

$$\int_{-\pi}^{\pi}\cos nx\cos mx\,dx=\frac{1}{2}\int_{-\pi}^{\pi}[\cos(m-n)x+\cos(m+n)x]\,dx=$$
$$=\frac{1}{2}\left[\frac{\sin(m-n)x}{m-n}+\frac{\sin(m+n)x}{m+n}\right]_{-\pi}^{\pi}=0 \tag{93}$$

(при $m \neq n$) и аналогично

$$\int_{-\pi}^{\pi}\sin nx\sin mx\,dx=0 \qquad (\text{при } m\neq n),$$

$$\int_{-\pi}^{\pi}\cos nx\sin mx\,dx=0 \qquad (\text{при любых } m,\ n=0,\ 1,\ 2,\ \ldots).$$

Система (92) ортогональна также на интервале $0 \leqslant x \leqslant 2\pi$ и вообще на любом интервале длины 2π. Это вытекает из свойства 10 п. XIV.4, если взять в качестве $f(x)$ произведение двух функций (92), а $A = 2\pi$.

Если воспользоваться свойством интеграла от четной функции (п. XIV.4, свойство 9), то из (93) получим

$$\int_{-\pi}^{\pi} \cos nx \cos mx\, dx = 2\int_{0}^{\pi} \cos nx \cos mx\, dx = 0$$

$$(m,\ n = 0,\ 1,\ 2,\ \ldots;\ m \neq n),$$

т. е. система функций

$$1,\ \cos x,\ \cos 2x,\ \ldots,\ \cos nx,\ \ldots \tag{94}$$

ортогональна на интервале $0 \leqslant x \leqslant \pi$. Аналогично проверяется, что на том же интервале система функций

$$\sin x,\ \sin 2x,\ \ldots,\ \sin nx,\ \ldots \tag{95}$$

также ортогональна. (Проверьте, что на интервале $0 \leqslant x \leqslant \pi$ система функций (92) не ортогональна.)

Если «растянуть» вдоль оси x каждую из функций (92) в $\frac{l}{\pi}$ раз, мы получим систему функций

$$1,\ \cos\frac{\pi x}{l},\ \sin\frac{\pi x}{l},\ \cos\frac{2\pi x}{l},\ \sin\frac{2\pi x}{l},\ \ldots,\ \cos\frac{n\pi x}{l},\ \sin\frac{n\pi x}{l},\ \ldots, \tag{96}$$

ортогональную на интервале $-l \leqslant x \leqslant l$. Подобным образом можно равномерно «растянуть» системы функций (94) и (95) и вообще любую ортогональную систему функций. Применяется также сдвиг ортогональной системы функций вдоль оси абсцисс, отчего она не перестает быть ортогональной (на сдвинутом интервале).

Свойством ортогональности могут обладать не только тригонометрические функции. Например, сейчас мы построим систему ортогональных полиномов на интервале $-1 \leqslant x \leqslant 1$.

Будем исходить из системы функций

$$1,\ x,\ x^2,\ x^3,\ \ldots,\ x^n,\ \ldots\quad (-1 \leqslant x \leqslant 1). \tag{97}$$

Первые две функции ортогональны друг другу:

$$\int_{-1}^{1} 1\cdot x\, dx = \frac{x^2}{2}\Big|_{-1}^{1} = 0,$$

поэтому положим $P_0(x) \equiv 1$, $P_1(x) \equiv x$. В то же время третья функция не ортогональна первой (проверьте!). Поэтому в качестве $P_2(x)$ возьмем линейную комбинацию первых трех функций (97), т. е. $P_2(x) = ax^2 + bx + c$, причем

коэффициенты a, b, c подберем так, чтобы $P_2(x)$ был ортогональным к уже построенным полиномам $P_0(x)$ и $P_1(x)$, т. е.

$$\int_{-1}^{1}(ax^2+bx+c)\cdot 1\cdot dx=0,\quad \int_{-1}^{1}(ax^2+bx+c)\cdot x\cdot dx=0.$$

Отсюда найдем (проверьте!)

$$b=0,\quad a=-3c,\quad \text{т. е.}\quad P_2(x)=c(-3x^2+1).$$

Здесь c — произвольная постоянная. Обычно она выбирается так, чтобы $P_2(1)=1$. (Такой выбор из равноправных объектов какого-либо одного называется *нормировкой*.) Получим $c=-\frac{1}{2}$, т. е. окончательно

$$P_2(x)=\frac{3}{2}x^2-\frac{1}{2}.$$

Для построения $P_3(x)$ берем комбинацию первых четырех функций (88), т. е. $P_3(x)=ax^3+bx^2+cx+d$, причем коэффициенты a, b, c, d подбираем так, чтобы $P_3(x)$ был ортогональным к уже построенным полиномам $P_0(x)$, $P_1(x)$ и $P_2(x)$. Отсюда при дополнительном условии $P_3(1)=1$ найдем подобно предыдущему (проверьте!)

$$P_3(x)=\frac{5}{2}x^3-\frac{3}{2}x.$$

Аналогично строим

$$P_4(x)=\frac{1}{8}(35x^4-30x^2+3),\qquad P_5(x)=\frac{1}{8}(63x^5-70x^3+15x)\ \text{и т. д.}$$

Эти многочлены ортогональны друг другу на интервале $-1\leqslant x\leqslant 1$. Они были исследованы Лежандром в 1783—1785 гг. и их называют теперь *многочленами Лежандра*. Они играют важную роль в различных областях математики и физики.

Подобный процесс *ортогонализации*, который мы проделали над системой функций (97) на интервале $-1\leqslant x\leqslant 1$, можно проделать над *любой* линейно независимой системой функций на *любом* интервале, если интегралы от квадратов этих функций по рассматриваемому интервалу сходятся.

21. Ряды по ортогональным функциям. Пусть задана система функций

$$g_1(x),\ g_2(x),\ \ldots,\ g_n(x),\ \ldots, \tag{98}$$

ортогональная на некотором интервале $a<x<b$. Может возникнуть задача о разложении произвольной функции $f(x)$ на этом интервале в ряд по функциям (98), т. е. в ряд вида

$$f(x)=a_1g_1(x)+a_2g_2(x)+\ldots+a_ng_n(x)+\ldots=\sum_{n=1}^{\infty}a_ng_n(x), \tag{99}$$

где a_n — числовые коэффициенты. При этом возникают вопросы: возможно ли разложение для любой функции $f(x)$? как найти коэффициенты a_n? как сходится ряд (99)?

Будем считать для простоты все рассматриваемые функции, а также интервал $a < x < b$ конечными. Ответ на первый вопрос зависит от выбора системы (98). Если разложение (99) возможно для любой функции $f(x)$, то система функций (98) называется *полной*. Можно доказать, что *все приведенные в п.* 20 *ортогональные системы функций являются полными* на указанных там интервалах.

Перейдем теперь к нахождению коэффициентов a_n разложения (99), причем будем считать, что ни одна из функций (98) не равна тождественно нулю. Для этого умножим обе части (99) на $g_n(x)$ и проинтегрируем результат по интервалу $a \leqslant x \leqslant b$:

$$\int_a^b f(x)\, g_n(x)\, dx =$$

$$= a_1 \int_a^b g_1(x)\, g_n(x)\, dx + a_2 \int_a^b g_2(x)\, g_n(x)\, dx + \ldots + a_n \int_a^b g_n^2(x)\, dx + \ldots$$

В силу ортогональности системы (98), в правой части последнего равенства все интегралы равны нулю, за исключением интеграла от $g_n^2(x)$, и мы получаем формулу для коэффициентов

$$a_n = \frac{\int_a^b f(x)\, g_n(x)\, dx}{\int_a^b g_n^2(x)\, dx} \qquad (n = 1,\ 2,\ 3,\ \ldots). \tag{100}$$

Так как коэффициенты находятся однозначно, то мы заключаем в частности, что *если суммы двух рядов* (99) *тождественно равны то и коэффициенты при одинаковых функциях* $g_n(x)$ *в этих рядах также одинаковы*; *если сумма ряда* (99) *тождественно равна нулю, то и все коэффициенты равны нулю.*

22. Ряды Фурье. Эти общие результаты можно применить к конкретным ортогональным системам функций. Так, для системы (92) получаем, что любую конечную функцию, заданную при $-\pi \leqslant x \leqslant \pi$, можно разложить в ряд вида

$$f(x) = a_1 + a_2 \cos x + a_3 \sin x + a_4 \cos 2x + a_5 \sin 2x + \ldots$$

Здесь удобнее обозначить коэффициенты по-другому и написать

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ldots =$$

$$= a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx). \tag{101}$$

Коэффициенты этого ряда находятся по формулам (100):

$$\left.\begin{aligned}
a_0 &= \frac{\int\limits_{-\pi}^{\pi} f(x)\cdot 1\,dx}{\int\limits_{-\pi}^{\pi} 1^2 dx} = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} f(x)\,dx,\\
a_n &= \frac{\int\limits_{-\pi}^{\pi} f(x)\cos nx\,dx}{\int\limits_{-\pi}^{\pi} \cos^2 nx\,dx} = \frac{1}{\pi}\int\limits_{-\pi}^{\pi} f(x)\cos nx\,dx,\\
b_n &= \frac{\int\limits_{-\pi}^{\pi} f(x)\sin nx\,dx}{\int\limits_{-\pi}^{\pi} \sin^2 nx\,dx} = \frac{1}{\pi}\int\limits_{-\pi}^{\pi} f(x)\sin nx\,dx \quad (n \geqslant 1).
\end{aligned}\right\} \qquad (102)$$

Аналогично получаются ряды по системам функций (94) или (95)

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos nx \qquad (0 \leqslant x \leqslant \pi), \qquad (103)$$

$$a_0 = \frac{1}{\pi}\int\limits_0^{\pi} f(x)\,dx, \quad a_n = \frac{2}{\pi}\int\limits_0^{\pi} f(x)\cos nx\,dx \qquad (n \geqslant 1);$$

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx \quad (0 \leqslant x \leqslant \pi), \quad b_n = \frac{1}{\pi}\int\limits_0^{\pi} f(x)\sin nx\,dx. \qquad (104)$$

Часто применяются ряды по функциям (96) и по подобным же образом растянутым системам функций (94) или (95):

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{\pi n x}{l} + b_n \sin\frac{\pi n x}{l}\right) \qquad (-l \leqslant x \leqslant l), \qquad (105)$$

$$a_0 = \frac{1}{2l}\int\limits_{-l}^{l} f(x)\,dx, \quad a_n = \frac{1}{l}\int\limits_{-l}^{l} f(x)\cos\frac{\pi n x}{l}\,dx,$$

$$b_n = \frac{1}{l}\int\limits_{-l}^{l} f(x)\sin\frac{\pi n x}{l}\,dx \qquad (n \geqslant 1);$$

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos\frac{\pi n x}{l} \qquad (0 \leqslant x \leqslant l), \qquad (106)$$

$$a_0=\frac{1}{l}\int_0^l f(x)\,dx, \quad a_n=\frac{2}{l}\int_0^l f(x)\cos\frac{\pi n x}{l}\,dx \qquad (n\geqslant 1);$$

$$f(x)=\sum_{n=1}^{\infty} b_n \sin\frac{\pi n x}{l} \qquad (0\leqslant x\leqslant l), \tag{107}$$

$$b_n=\frac{2}{l}\int_0^l f(x)\sin\frac{\pi n x}{l}\,dx \qquad (n\geqslant 1).$$

Ряды (101), (103) и (104) являются частными случаями рядов (105), (106) и (107), именно, первые получаются из вторых при $l=\pi$. Все они называются *рядами Фурье* по имени выдающегося французского математика Ж. Фурье́ (1768—1830), который в 1811 г. впервые их широко использовал в своих исследованиях по теплопроводности, хотя такие ряды иногда применялись и ранее. Формулы (102) для *коэффициентов Фурье* и другие подобные формулы были получены еще Эйлером в 1777 г.

Приведем примеры разложения в ряд Фурье. Пусть функцию $y=x$ требуется разложить при $0\leqslant x\leqslant l$ в ряд (106); для этого подсчитываем коэффициенты:

$$a_0=\frac{1}{l}\int_0^l x\,dx=\frac{l}{2};$$

$$a_n=\frac{2}{l}\int_0^l x\cos\frac{\pi n x}{l}\,dx=\frac{2}{l}\,x\frac{l}{\pi n}\sin\frac{\pi n x}{l}\Big|_0^l-\frac{2}{l}\int_0^l \frac{l}{\pi n}\sin\frac{\pi n x}{l}\,dx=$$

$$=\frac{2}{\pi n}\frac{l}{\pi n}\cos\frac{\pi n x}{l}\Big|_0^l=\frac{2l}{\pi^2 n^2}(\cos\pi n-1)=-\frac{2l}{\pi^2 n^2}[1-(-1)^n] \qquad (n\geqslant 1).$$

Отсюда получаем

$$a_1=-\frac{4l}{\pi^2 1^2}, \quad a_2=0,\ a_3=-\frac{4l}{\pi^2 3^2}, \quad a_4=0,\ \ldots$$

и окончательно

$$x=l\left[\frac{1}{2}-\frac{4}{\pi^2}\left(\frac{1}{1^2}\cos\frac{\pi x}{l}+\frac{1}{3^2}\cos\frac{3\pi x}{l}+\frac{1}{5^2}\cos\frac{5\pi x}{l}+\ldots\right)\right] (0\leqslant x\leqslant l). \tag{108}$$

В качестве второго примера возьмем функцию, *заданную несколькими формулами* (график ее составляет среднюю часть рис. 340, *б*):

$$f(x)=1\ (-l<x<-l+\alpha), \quad =0\ (-l+\alpha<x<0),$$
$$=1\ (0<x<\alpha), \quad =0\ (\alpha<x<l),$$

где α — некоторое число, $0<\alpha<l$, и разложим ее в ряд (105):

$$a_0=\frac{1}{2l}\int_{-l}^{l}f(x)\,dx=\frac{1}{2l}\left(\int_{-l}^{-l+\alpha}+\int_{-l+\alpha}^{0}+\int_{0}^{\alpha}+\int_{\alpha}^{l}\right)f(x)\,dx=$$

$$=\frac{1}{2l}\left(\int_{-l}^{-l+\alpha}1\,dx+\int_{-l+\alpha}^{0}0\,dx+\int_{0}^{\alpha}1\,dx+\int_{\alpha}^{l}0\,dx\right)=\frac{1}{2l}(\alpha+0+\alpha+0)=\frac{\alpha}{l}.$$

(Аналогичным образом находят интеграл от любой функции, заданной

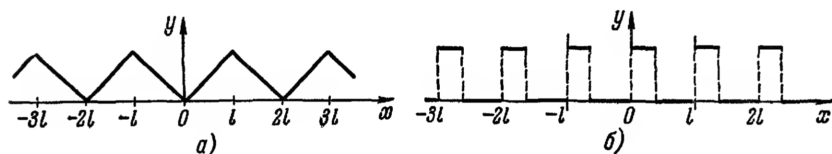


Рис. 340.

несколькими формулами.) Далее,

$$a_n=\frac{1}{l}\int_{-l}^{l}f(x)\cos\frac{k\pi x}{l}\,dx=\frac{1}{l}\left(\int_{-l}^{-l+\alpha}\cos\frac{n\pi x}{l}\,dx+\int_{0}^{\alpha}\cos\frac{n\pi x}{l}\,dx\right)=$$

$$=\frac{1}{n\pi}\left[\sin\frac{n\pi(-l+\alpha)}{l}+\sin\frac{n\pi\alpha}{l}\right]=$$

$$=\frac{1}{n\pi}\left(-\sin\frac{n\pi l}{l}\cos\frac{n\pi\alpha}{l}+\cos\frac{n\pi l}{l}\sin\frac{n\pi\alpha}{l}+\sin\frac{n\pi\alpha}{l}\right)=$$

$$=\frac{1}{n\pi}[(-1)^n+1]\sin\frac{n\pi\alpha}{l},$$

$$b_n=\frac{1}{l}\left(\int_{-l}^{-l+\alpha}\sin\frac{n\pi x}{l}\,dx+\int_{0}^{\alpha}\sin\frac{n\pi x}{l}\,dx\right)=$$

$$=-\frac{1}{n\pi}\left[\cos\frac{n\pi(-l+\alpha)}{l}-\cos(-n\pi)+\cos\frac{n\pi\alpha}{l}-1\right]=$$

$$=-\frac{1}{n\pi}\left[\cos\frac{n\pi l}{l}\cos\frac{n\pi\alpha}{l}+\sin\frac{n\pi l}{l}\sin\frac{n\pi\alpha}{l}-(-1)^n+\cos\frac{n\pi\alpha}{l}-1\right]=$$

$$=-\frac{1}{n\pi}\left\{[(-1)^n+1]\cos\frac{n\pi\alpha}{l}-(-1)^n-1\right\}=\frac{1}{n\pi}[(-1)^n+1]\left(1-\cos\frac{n\pi\alpha}{l}\right).$$

Таким образом, при n нечетном все $a_n=b_n=0$, тогда как

$$a_{2k}=\frac{1}{k\pi}\sin\frac{2k\pi\alpha}{l},\quad b_{2k}=\frac{1}{k\pi}\left(1-\cos\frac{2k\pi\alpha}{l}\right)\qquad(k=1,\ 2,\ 3,\ \ldots).$$

Особенно простой результат получается, если $\alpha=\frac{l}{2}$; тогда

$$a_{2k}=\frac{1}{k\pi}\sin k\pi=0,\qquad b_{2k}=\frac{1}{k\pi}(1-\cos k\pi)=\frac{1}{k\pi}[1-(-1)^k]$$

и ряд принимает вид

$$f(x)=\frac{1}{2}+\frac{2}{\pi}\left(\frac{1}{1}\sin\frac{2\pi x}{l}+\frac{1}{3}\sin\frac{6\pi x}{l}+\frac{1}{5}\sin\frac{10\pi x}{l}+\ldots\right). \qquad (109)$$

Таким образом, функцию, заданную несколькими формулами, оказывается возможным представить в виде единого ряда. Открытие Фурье этого факта сыграло в истории математики большую роль, так как привело к значительному расширению содержания понятия функции.

При практическом разложении функции (особенно, заданной таблично или графически) в ряд Фурье обычно для подсчета коэффициентов пользуются формулами численного интегрирования (п. XIV.13). Пусть, например, рассматривается разложение в ряд (107) и мы хотим воспользоваться формулой трапеций, разбив отрезок интегрирования на 24 части. Тогда, обозначив $x_k=\frac{kl}{24}$, $f_k=f(x_k)$ $(k=0,\ 1,\ \ldots,\ 24)$, получим

$$b_n=\frac{2}{l}\int_0^l f(x)\sin\frac{n\pi x}{l}\,dx\approx$$

$$\approx\frac{2}{l}\,\frac{l}{24}\left(\frac{f_0}{2}\sin\frac{n\pi x_0}{l}+f_1\sin\frac{n\pi x_1}{l}+\ldots+\frac{f_{24}}{2}\sin\frac{n\pi x_{24}}{l}\right)=$$

$$=\frac{1}{12}\left(\frac{f_0}{2}\sin 0^\circ n+f_1\sin 7{,}5^\circ n+f_2\sin 15^\circ n+\ldots+\frac{f_{24}}{2}\sin 180^\circ n\right). \qquad (110)$$

Нетрудно видеть, что здесь при любом n требуются только следующие значения синуса:

$$\begin{array}{rr}
\sin 0^\circ=0{,}0000, & \sin 52{,}5^\circ=0{,}7934,\\
\sin 7{,}5^\circ=0{,}1305, & \sin 60^\circ=0{,}8660,\\
\sin 15^\circ=0{,}2588, & \sin 67{,}5^\circ=0{,}9239,\\
\sin 22{,}5^\circ=0{,}3827, & \sin 75^\circ=0{,}9659,\\
\sin 30^\circ=0{,}5000, & \sin 82{,}5^\circ=0{,}9914,\\
\sin 37{,}5^\circ=0{,}6088, & \sin 90^\circ=1{,}0000.\\
\sin 45^\circ=0{,}7071, &
\end{array}$$

При применении формулы (110) для данного n нужно подставить значения синуса из этой таблицы, пользуясь тригонометрическими формулами приведения; затем сгруппировать члены с одинаковыми вторыми множителями, сложить в этих членах значения f_k и, после умножения, подсчитать полную сумму.

23. Разложение периодической функции. Ряды Фурье применяются не только для разложения функции, заданной на конечном интервале, но и для разложения периодической функции, заданной на всей оси. Предположим сначала, что функция $f(x)$ задана при $-\pi\leqslant x\leqslant\pi$, и разложим ее в ряд (101). Его члены имеют смысл и вне данного интервала, причем их период равен 2π (п. I.16), так как

$$\cos n(x+2\pi)\equiv\cos(nx+2\pi n)\equiv\cos nx;$$
$$\sin n(x+2\pi)\equiv\sin nx \qquad (n=1,\ 2,\ 3,\ \ldots).$$

Значит, и вся сумма периодична с периодом 2π. Но так как на интервале $-\pi \leqslant x \leqslant \pi$ она равна $f(x)$, то *сумма ряда* (101) *дает периодическое продолжение функции* $f(x)$ *с интервала* $-\pi \leqslant x \leqslant \pi$ *на всю ось* x *с периодом* 2π.

Аналогично, так как все члены ряда (103) — четные функции, то его сумма дает четное периодическое продолжение функции $f(x)$ с интервала $0 \leqslant x \leqslant \pi$ на всю ось с периодом 2π; сумма же ряда (104) дает нечетное продолжение с тем же периодом. Аналогичный результат получается для рядов (105)—(107), где, однако, период равен $2l$.

Так, на рис. 340 изображены рассмотренные на всей оси x суммы рядов, разобранных в примерах п. 22; отметим, что во втором примере, хотя $2l$ и являются периодом, наименьший период равен l.

Пусть теперь функция $f(x)$ с самого начала задана *на всей оси* x *как периодическая функция с периодом* 2π. Если тогда составить ряд (101), где коэффициенты вычислены по формулам (102), то он, как мы показали, дает периодическое продолжение $f(x)$ с отрезка $-\pi \leqslant x \leqslant \pi$ на всю ось с периодом 2π, т. е. совпадает с $f(x)$ на всей оси x.

При этом коэффициенты можно вычислять также и по формулам $a_0 = \frac{1}{2\pi} \int\limits_{\alpha-\pi}^{\alpha+\pi} f(x)\,dx$ (α любое); в частности, $a_0 = \frac{1}{2\pi} \int\limits_0^{2\pi} f(x)\,dx$, и аналогично для коэффициентов a_n, b_n; это вытекает из периодичности подынтегральной функции (свойство 10 п. XIV.4).

Подобным образом четная (нечетная) периодическая функция с периодом 2π разлагается в ряд (103) (соответственно (104)); для периода $2l$ получаются ряды (105)—(107).

Разложение (105) часто преобразуют с помощью формулы (I.18). Получится

$$f(x) = a_0 + \sum_{n=1}^{\infty} M_n \sin\left(\frac{\pi n x}{l} + \alpha_n\right). \tag{111}$$

Постоянная a_0 равна среднему значению функции $f(x)$ на всей оси (п. XIV.5), так как средние значения прочих слагаемых равны нулю (проверьте!). Первое непостоянное слагаемое называется *основной гармоникой*; оно имеет период как раз $2l$. Дальнейшие слагаемые называются *верхними гармониками*; их наименьшие периоды равны последовательно $\frac{2l}{2}$, $\frac{2l}{3}$, $\frac{2l}{4}$ и т. д. Поэтому само разложение периодической функции в ряд Фурье, т. е. в сумму гармоник, называется также *гармоническим анализом*.

Если независимая переменная истолковывается как время, то лучше период обозначить через T, а формулу (111) переписать в виде

$$f(t) = a_0 + \sum_{n=1}^{\infty} M_n \sin(n\omega t + \alpha_n) \qquad \left(\omega = \frac{2\pi}{T}\right).$$

Таким образом, ряд Фурье осуществляет представление произвольного периодического колебания в виде суммы гармонических колебаний с кратными частотами. Это разложение, в частности, хорошо известно в акустике, где основное слагаемое $M_1 \sin(\omega t+\alpha_1)$ определяет высоту звука, *основной тон*, а дальнейшим слагаемым соответствуют *обертоны*, от которых зависит *тембр* звука.

В разложении (112) все периоды слагаемых *соизмеримы* между собой, т. е. их отношения равны рациональным числам. Это связано с тем общим фактом, что, как можно доказать, *сумма периодических функций с различными периодами будет периодической в том и только в том случае, если периоды у слагаемых соизмеримые*. Суммы периодических функций с несоизмеримыми периодами принадлежат к более широкому классу *почти периодических функций*, которые сейчас широко применяются, в частности, при рассмотрении результата наложения несинхронизированных колебаний.

24. Пример. Функции Бесселя как коэффициенты Фурье. В теоретической радиотехнике рассматривается функция

$$e^{ix\cos t}=\cos(x\cos t)+i\sin(x\cos t). \tag{112}$$

При фиксированном x она периодична по t с периодом 2π и четна, а потому разлагается в ряд Фурье вида (103). Чтобы получить это разложение, можно перемножить ряды

$$e^{\frac{ix}{2}e^{it}}=1+\frac{1}{1!}\left(\frac{ix}{2}\right)e^{it}+\frac{1}{2!}\left(\frac{ix}{2}\right)^2 e^{i2t}+\frac{1}{3!}\left(\frac{ix}{2}\right)^3 e^{i3t}+\ldots,$$

$$e^{\frac{ix}{2}e^{-it}}=1+\frac{1}{1!}\left(\frac{ix}{2}\right)e^{-it}+\frac{1}{2!}\left(\frac{ix}{2}\right)^2 e^{-i2t}+\frac{1}{3!}\left(\frac{ix}{2}\right)^3 e^{-i3t}+\ldots.$$

После умножения слева получится функция (112), а справа объединим члены с одинаковыми экспонентами. Именно, при e^{ikt} и при e^{-ikt} получится коэффициент

$$1\cdot\frac{1}{k!}\left(\frac{ix}{2}\right)^k+\frac{1}{1!}\left(\frac{ix}{2}\right)\frac{1}{(k+1)!}\left(\frac{ix}{2}\right)^{k+1}+\frac{1}{2!}\left(\frac{ix}{2}\right)^2\frac{1}{(k+2)!}\left(\frac{ix}{2}\right)^{k+2}+\ldots=$$
$$=i^k\left(\frac{1}{0!\,k!}\left(\frac{x}{2}\right)^k-\frac{1}{1!\,(k+1)!}\left(\frac{x}{2}\right)^{k+2}+\frac{1}{2!\,(k+2)!}\left(\frac{x}{2}\right)^{k+1}-\ldots\right)=i^kJ_k(x)$$

(п. XV.26). Итак,

$$e^{ix\cos t}=J_0(x)+\sum_{k=1}^{\infty}i^kJ_k(x)\left(e^{ikt}+e^{-ikt}\right)=J_0(x)+2\sum_{k=1}^{\infty}i^kJ_k(x)\cos kt. \tag{113}$$

Это и есть требуемое разложение в ряд Фурье. Отделяя в нем вещественную часть от мнимой, получим также разложения

$$\cos(x\cos t)=J_0(x)-2J_2(x)\cos 2t+2J_4(x)\cos 4t-\ldots,$$
$$\sin(x\cos t)=2J_1(x)\cos t-2J_3(x)\cos 3t+2J_5(x)\cos 5t-\ldots$$

Из формулы (113) вытекает, в частности, *интегральное представление функций Бесселя целого порядка* (проверьте на основе формул для коэффициентов ряда (103)):

$$J_n(x)=\frac{1}{\pi i^n}\int_0^{\pi}e^{ix\cos t}\cos nt\,dt.$$

25. Характер сходимости ряда Фурье. Пусть ограниченная периодическая функция $f(x)$ периода $2l$ разложена в ряд Фурье (105). Легко проверить, что *все коэффициенты Фурье ограничены*; например, оценим коэффициенты a_n:

$$|a_n| = \frac{1}{l}\left|\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx\right| \leqslant \frac{1}{l}\int_{-l}^{l} |f(x)|\left|\cos\frac{n\pi x}{l}\right|dx \leqslant$$

$$\leqslant \frac{1}{l}\int_{-l}^{l} |f(x)|\,dx.$$

Тот же результат получится, если функция $f(x)$ не ограничена, но абсолютно интегрируема на своем периоде, т. е. если

$$\int_{-l}^{l} |f(x)|\,dx < \infty.$$

Пусть функция $f(x)$ ограничена и разрывна. Тогда ее коэффициенты Фурье *убывают со скоростью* $\frac{1}{n}$, как во втором примере п. 22.

В самом деле, допустим для определенности, что $f(x)$ имеет на своем периоде два разрыва, при $x=x_1$ и $x=x_2$, причем $-l<x_1<x_2<l$. Тогда

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx =$$

$$= \frac{1}{l}\int_{-l}^{x_1} f(x)\cos\frac{n\pi x}{l}\,dx + \frac{1}{l}\int_{x_1}^{x_2} f(x)\cos\frac{n\pi x}{l}\,dx + \frac{1}{l}\int_{x_2}^{l} f(x)\cos\frac{n\pi x}{l}\,dx.$$

Каждый из этих интегралов проинтегрируем по частям:

$$a_n = \frac{1}{n\pi}\left[f(x_1-0)\sin\frac{n\pi x_1}{l} - f(-l)\sin\frac{n\pi(-l)}{l} - \int_{-l}^{x_1} f'(x)\sin\frac{n\pi x}{l}\,dx\right] + \ldots =$$

$$= -\frac{1}{n\pi}[f(x_1+0)-f(x_1-0)]\sin\frac{n\pi x_1}{l} - \frac{1}{n\pi}[f(x_2+0)-$$

$$-f(x_2-0)]\sin\frac{n\pi x_2}{l} - \frac{1}{n\pi}\int_{-l}^{l} f'(x)\sin\frac{n\pi x}{l}\,dx. \quad (114)$$

Последний интеграл отличается от коэффициента Фурье для $f'(x)$ лишь постоянным множителем l. Однако $\int_{-l}^{l} |f'(x)|\,dx < \infty$, так как на каждом интервале (α, β) монотонности и непрерывности функции $f(x)$ производная $f'(x)$ сохраняет знак и потому

$$\int_{\alpha}^{\beta} |f'(x)|\,dx = \left|\int_{\alpha}^{\beta} f'(x)\,dx\right| = |f(\beta-0) - f(\alpha+0)| < \infty.$$

(При этом мы исключаем функции, имеющие бесконечное число интервалов монотонности на конечном интервале изменения x (рис. 102), как чрезвычайно редкие.) Согласно предыдущему последний интеграл в формуле (114) ограничен, откуда и следует наше утверждение о скорости убывания коэффициентов Фурье.

Пусть теперь сама функция $f(x)$ непрерывна, но ее производная имеет разрывы и ограничена, как в первом примере п. 22. Тогда коэффициенты Фурье *убывают со скоростью* $\frac{1}{n^2}$.

В самом деле, интегрируя по частям выражение для коэффициента Фурье, получим

$$a_n=\frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx=-\frac{1}{n\pi}\int_{-l}^{l} f'(x)\sin\frac{n\pi x}{l}\,dx.$$

Применяя к интегралу, стоящему в правой части, рассуждения последних абзацев, получим, что при $n\to\infty$ он убывает со скоростью $\frac{1}{n}$, а значит, a_n и аналогично b_n убывают со скоростью $\frac{1}{n^2}$. При помощи двукратного интегрирования по частям доказывается, что если $f(x)$ и $f'(x)$ непрерывны, а $f''(x)$ имеет разрывы, то коэффициенты убывают как $\frac{1}{n^3}$ и т. д. Итак, *чем функция «глаже»*, т. е. чем больше она имеет непрерывных производных, тем быстрее убывают коэффициенты, т. е. *тем быстрее сходится ее ряд Фурье.*

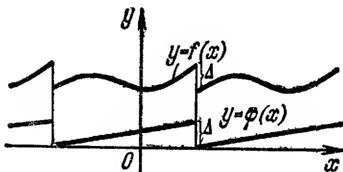


Рис. 341.

Ряд Фурье для разрывной функции $f(x)$ сходится очень медленно и вычисления с его помощью практически затруднительны. Поэтому иногда поступают так: подбирают какую-либо по возможности более простую функцию $\varphi(x)$, имеющую разрывы в тех же точках, что и $f(x)$, причем с такими же скачками (рис. 341). Тогда разность $f(x)-\varphi(x)$ уже не имеет разрывов и потому разлагается в ряд Фурье, сходящийся лучше, чем ряд для $f(x)$. Поэтому $f(x)$ равно сумме простой функции $\varphi(x)$ и сравнительно неплохо сходящегося тригонометрического ряда. Аналогично можно устранить разрывы у первой производной и т. д. (Ср. метод выделения особенностей в п. 4.)

Итак, если функция $f(x)$ непрерывна, а ее производная ограничена, то коэффициенты Фурье убывают со скоростью не менее $\frac{1}{n^2}$. Но так как

$$\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty$$

(п. 1), то в силу признака Вейерштрасса (п. 8) ряд Фурье сходится равномерно на всей оси x. Более детальное исследование показывает, что при этом условие, наложенное нами на $f'(x)$, является излиш-

ним, так как для применимости признака Вейерштрасса коэффициенты могут стремиться к нулю с еще меньшей скоростью. В данном случае при подстановке в ряд численных значений x мы получим $f(x)$ тождественно.

Если функция $f(x)$ разрывна, то ее ряд Фурье заведомо не может сходиться равномерно, так как его члены — непрерывные функции (п. 9, свойство 1). Можно показать, что в этом случае при подстановке в ряд численных значений x получается $f(x)$ во всех точках непрерывности функции f и

$$\frac{f(x-0)+f(x+0)}{2}$$

в ее точках разрыва. Так, ряд (109) при $x=0$ имеет сумму $\frac{1}{2}$, так как в данном случае $f(-0)=0$, $f(+0)=1$.

Если функцию разлагать в ряд Фурье на конечном интервале (п. 22), то скорость сходимости ряда определяется разрывами этой функции и ее производных *после ее продолжения на всю ось* x, описанного в п. 23. Так, в первом примере п. 22 коэффициенты имели порядок $\frac{1}{n^2}$, так как после продолжения (рис. 340, a) получилась функция с разрывами у первой производной. Если ту же функцию $y=x$ раскладывать в ряд по синусам (ряд (107)), то коэффициенты имели бы порядок $\frac{1}{n}$, так как тогда после продолжения имела бы разрывы сама функция (почему?).

С помощью рядов Фурье можно найти суммы многих интересных числовых рядов. Так, если в ряд (109) подставить $x=\frac{l}{4}$, то получится

$$1=\frac{1}{2}+\frac{2}{\pi}\left(\frac{1}{1}\sin\frac{\pi}{2}+\frac{1}{3}\sin\frac{3\pi}{2}+\frac{1}{5}\sin\frac{5\pi}{2}+\ldots\right)=$$
$$=\frac{1}{2}+\frac{2}{\pi}\left(\frac{1}{1}-\frac{1}{3}+\frac{1}{5}-\ldots\right),$$

откуда

$$\frac{1}{1}-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\ldots=\frac{\pi}{4}.$$

Если в ряд (108) подставить $x=0$, то получится

$$0=l\left[\frac{1}{2}-\frac{4}{\pi^2}\left(\frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\ldots\right)\right],$$

т. е.

$$\frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\ldots=\frac{\pi^2}{8}.$$

Отсюда в свою очередь можно найти значение $\zeta(2)$ (формула (19)):

$$\zeta(2)=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\ldots=$$
$$=\frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{5^2}+\ldots+\frac{1}{2^2}\left(\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\ldots\right)=\frac{\pi^2}{8}+\frac{\zeta(2)}{4},$$

откуда

$$\zeta(2)=\frac{\pi^2}{6}=1,645.$$

26. Комплексная форма ряда Фурье. В ряде Фурье можно с помощью формулы Эйлера перейти от тригонометрических функций к показательным, что иногда бывает предпочтительнее. Например, если проделать такое преобразование над рядом (105), то надо воспользоваться формулами

$$\cos\frac{\pi nx}{l}=\frac{e^{\frac{i\pi nx}{l}}+e^{-\frac{i\pi nx}{l}}}{2},$$

$$\sin\frac{\pi nx}{2}=\frac{e^{\frac{i\pi nx}{l}}-e^{-\frac{i\pi nx}{l}}}{2i}.$$

После подстановки и приведения подобных членов в правой части получится сумма ряда по всем этим экспонентам, другими словами, по экспонентам $e^{\frac{i\pi nx}{l}}$, где $n=0,\ \pm 1,\ \pm 2,\ \ldots$ Итак,

$$f(x)=\sum_{n=-\infty}^{\infty} c_n e^{\frac{i\pi nx}{l}}. \tag{115}$$

Чтобы найти коэффициенты c_n, умножим обе части на $e^{-\frac{i\pi nx}{l}}$ при каком-либо фиксированном n и проинтегрируем результат от $-l$ до l. Для членов, номер которых не равен выбранному n, интеграл равен

$$\int_{-l}^{l} c_m e^{\frac{i\pi mx}{l}} e^{-\frac{i\pi nx}{l}}\,dx=c_m \frac{e^{\frac{i\pi(m-n)}{l}}}{i\pi(m-n)}\,l\,\Big|_{-l}^{l}=$$

$$=\frac{c_m l}{i\pi(m-n)}\left[e^{i\pi(m-n)}-e^{-i\pi(m-n)}\right]=\frac{2c_m l}{\pi(m-n)}\sin\pi(m-n)=0,$$

тогда как для члена с номером n интеграл равен $2lc_n$. Итак, коэффициенты ряда (115) вычисляются по формуле

$$c_n=\frac{1}{2l}\int_{-l}^{l} f(x)e^{-\frac{i\pi nx}{l}}\,dx. \tag{116}$$

В общем случае разложение по комплексным ортогональным функциям происходит так. Две *комплексные функции* $g(x)$ и $h(x)$ называются *ортогональными* друг другу на отрезке $a\leqslant x\leqslant b$, если

$$\int_a^b g(x)\,h^*(x)\,dx=0, \tag{117}$$

где звездочкой обозначена сопряженная функция (п. VIII.3). В частном случае, когда g и h вещественные, это определение переходит в определение (91). Отметим, что если взять от обеих частей (117) сопряженные величины, то мы получим

$$\left(\int_a^b g(x)\,h^*(x)\,dx\right)^* = \int_a^b (g(x)\,h^*(x))^*\,dx = \int_a^b h(x)\,g^*(x)\,dx = 0,$$

т. е. условие ортогональности является взаимным: если g ортогональна h, то h ортогональна g.

Если имеется система комплексных функций

$$g_1(x),\ g_2(x),\ \ldots,\ g_n(x),\ \ldots, \tag{118}$$

ортогональных на отрезке $a \leqslant x \leqslant b$, и если какую-либо комплексную функцию $f(x)$ можно разложить в ряд по этим функциям (так будет всегда, если система (108) полная), то

$$f(x) = c_1 g_1(x) + c_2 g_2(x) + \ldots + c_n g_n(x) + \ldots = \sum_{n=1}^{\infty} c_n g_n(x). \tag{119}$$

Чтобы найти коэффициенты разложения, надо обе части умножить на $g_n^*(x)$ при каком-либо фиксированном n и проинтегрировать результат от a до b. В силу условия ортогональности в правой части останется только один член и мы получим формулу

$$c_n = \frac{\int_a^b f(x)\,g_n^*(x)\,dx}{\int_a^b g_n(x)\,g_n^*(x)\,dx} = \frac{\int_a^b f(x)\,g_n^*(x)\,dx}{\int_a^b |g_n(x)|^2\,dx}.$$

В частности, разложение (115) — это разложение по ортогональной системе функций

$$\ldots,\ e^{-\frac{i2\pi x}{l}},\ e^{-\frac{i\pi x}{l}},\ 1,\ e^{\frac{i\pi x}{l}},\ e^{\frac{i2\pi x}{l}},\ \ldots,$$

полной на отрезке $-l \leqslant x \leqslant l$. Оно пригодно для любой ограниченной или даже неограниченной, но абсолютно интегрируемой (п. XIV.16) комплексной функции $f(x)$.

27. Равенство Парсеваля. Будем опять считать функции вещественными. Пусть система функций

$$g_1(x),\quad g_2(x),\ \ldots,\ g_n(x),\ \ldots \tag{120}$$

ортогональна и *полна* на интервале $a \leqslant x \leqslant b$. Рассмотрим разложение (99) произвольной функции $f(x)$ в ряд по этим функциям. Возводя обе части этого разложения в квадрат и интегрируя результат от a до b, получим в левой части

$$\int_a^b f^2(x)\,dx;$$

будем предполагать, что этот интеграл имеет конечное значение. В правой части после возведения в квадрат получится сумма квадратов отдельных членов и их попарных произведений. Интегралы от этих произведений равны нулю в силу ортогональности функций (120), а интервалы от квадратов

останутся, т. е.

$$\int_a^b f^2(x)\,dx = \sum_{n=1}^{\infty} a_n^2 \int_a^b g_n^2(x)\,dx. \tag{121}$$

Эта формула называется *равенством Парсевáля*. В частности, для ряда Фурье (101) получаем

$$\int_{-\pi}^{\pi} f^2(x)\,dx = 2\pi a_0^2 + \pi \sum_{n=1}^{\infty} (a_n^2 + b_n^2),$$

а для ряда (105)

$$\int_{-l}^{l} f^2(x)\,dx = 2la_0^2 + l \sum_{n=1}^{\infty} (a_n^2 + b_n^2).$$

Эти равенства были установлены в 1805 г. Из них между прочим следует, что при $n \to \infty$ будет $a_n \to 0$ и $b_n \to 0$.

Если система (120) не полна, то, как можно доказать, ее можно расширить до полной. Так как после такого расширения справедливо равенство вида (121), а все члены в правой части $\geqslant 0$, то до расширения имело место неравенство

$$\sum_{n=1}^{\infty} a_n^2 \int_a^b g_n^2(x)\,dx \leqslant \int_a^b f^2(x)\,dx;$$

при этом знак равенства здесь имеет место только для тех функций $f(x)$, которые можно разложить по системе функций (120).

Равенство Парсеваля (121) дает возможность по-новому подойти к построению ряда по ортогональным функциям. Пусть дано конечное число функций

$$g_1(x),\ g_2(x),\ \ldots,\ g_n(x), \tag{122}$$

ортогональных на отрезке $a \leqslant x \leqslant b$. Поставим задачу: для данной функции $f(x)$ составить линейную комбинацию функций (122) так, чтобы ее среднее квадратичное уклонение (п. 7) от $f(x)$ было минимальным.

Для решения этой задачи будем рассматривать функции (122) как часть полной ортогональной системы (120). Тогда

$$f(x) - \sum_{k=1}^{n} C_k g_k(x) = \sum_{k=1}^{\infty} a_k g_k(x) - \sum_{k=1}^{n} C_k g_k(x) =$$

$$= \sum_{k=1}^{n} (a_k - C_k)\, g_k(x) + \sum_{k=n+1}^{\infty} a_k g_k(x),$$

где a_k — коэффициенты разложения $f(x)$ по функциям (110), а C_k — произвольные постоянные. В силу равенства (121)

$$\int_a^b \Big[f(x) - \sum_{k=1}^{n} C_k g_k(x)\Big]^2 dx = \sum_{k=1}^{n} (a_k - C_k)^2 \int_a^b g_k^2(x)\,dx + \sum_{k=n+1}^{\infty} a_k^2 \int_a^b g_k^2(x)\,dx.$$

Если произвольно менять коэффициенты $C_1, C_2, \ldots, C_n$, то из последней формулы видно, что минимальное значение правой части получается, когда

$$C_1=a_1, \quad C_2=a_2, \ldots, C_n=a_n.$$

Итак, минимальное среднее квадратичное уклонение линейной комбинации функций (122) от $f(x)$ осуществляет отрезок разложения функции $f(x)$ в ряд по системе ортогональных функций, т. е. линейная комбинация, коэффициенты которой определены по формулам (100).

Аналогичное исследование полной системы комплексных ортогональных функций (118) и ряда (119) по ним приводит с учетом равенства $aa^*=|a|^2$ к формуле

$$\int_a^b |f(x)|^2\,dx=\sum_{n=1}^{\infty}|a_n|^2\int_a^b |g_n(x)|^2\,dx. \tag{123}$$

28. Пространство Гильберта. Вернемся к вещественным функциям, заданным на интервале $a \leqslant x \leqslant b$. Оказывается, что аналогия таких функций с векторами, упомянутая в п. 20, является весьма далекой. Так как над указанными функциями можно осуществлять линейные действия по обычным правилам, то эти функции образуют линейное пространство в смысле п. VII.17. Более того, если ввести понятие скалярного произведения двух функций по формуле

$$(f, g)=\int_a^b f(x)\,g(x)\,dx, \tag{124}$$

то нетрудно проверить, что все аксиомы п. VII.20 выполнены, так что здесь получается не только линейное, но даже евклидово пространство. К нему причисляются не только ограниченные, но также и неограниченные *квадратично суммируемые* функции, т. е. такие функции $f(x)$, что

$$(f, f)=\int_a^b f^2(x)\,dx<\infty. \tag{125}$$

Для таких функций из простого неравенства $2|f(x)\,g(x)| \leqslant f^2(x)+g^2(x)$ вытекает, что интеграл (124), если несобственный, то сходящийся.

Совокупность функций, удовлетворяющих условию (125), со скалярным произведением (124) называется *пространством Гильберта* L^2 по имени немецкого математика Д. Гильберта (1862—1943). Условие (91) в силу (124) и есть условие ортогональности векторов в этом пространстве. Полная ортогональная система функций — это ортогональный базис в пространстве L^2; отметим, что при определении полноты (п. 21) сходимость ряда (99) надо было понимать в смысле L^2, т. е. в смысле среднего квадратичного (п. 8). Формулы (100) для коэффициентов разложения — это те же формулы (VII.29), процесс ортогонализации, описанный в п. 20, — это тот же процесс, что был описан в п. VII.21. Равенство Парсеваля (121) — это аналог теоремы Пифагора в пространстве Гильберта: квадрат диагонали прямоугольного параллелепипеда равен сумме квадратов всех его измерений. (Попробуйте дать геометрическое истолкование доказанному в п. 27 минимальному свойству отрезка ряда (99).)

Характерной чертой пространства Гильберта является его *бесконечномерность*. Это, в частности, значительно затрудняет проверку полноты ортогональной системы функций. Система из k попарно ортогональных ненулевых векторов в n-мерном евклидовом пространстве является полной, если $k=n$, и неполной, если $k<n$. В отличие от этого в бесконечномерном пространстве

ортогональная система, состоящая из бесконечного числа функций, не обязана быть полной, так что здесь только по количеству функций в системе вопрос о ее полноте решить нельзя; этот вопрос сложный, и мы не будем его рассматривать.

Можно доказать, что всякая ортогональная неполная система представляет собой часть некоторой ортогональной полной системы. Поэтому в ряд по первой системе можно разложить те и только те функции, для которых при разложении по второй системе коэффициенты при дописанных функциях равны нулю.

В п. VII.20 мы показали, что в евклидовых пространствах имеет место важное неравенство (VII.26), которое теперь принимает вид, после возведения в квадрат,

$$\left(\int_a^b f(x)\,g(x)\,dx\right)^2 \leqslant \int_a^b f^2(x)\,dx \int_a^b g^2(x)\,dx.$$

Если в него вместо f и g подставить $|f|$ и 1, мы получим

$$\left(\int_a^b |f(x)|\,dx\right)^2 \leqslant (b-a)\int_a^b f^2(x)\,dx.$$

Отсюда видно, что все функции из L^2 являются суммируемыми, а также что из сходимости в среднем квадратичном вытекает сходимость в среднем (п. 8); обратное не обязательно. Впрочем, можно доказать, что ряд Фурье любой суммируемой функции сходится к ней в среднем.

29. Ортогональность с весом. При обычном интегрировании функции $f(x)$ по интервалу $a\leqslant x\leqslant b$ все значения x из этого интервала выступают как равноправные. Если хотят подчеркнуть бóльшую роль «вклада» одних значений по сравнению с другими, то рассматривают *интегрирование с весом*

$$\int_a^b f(x)\,\rho(x)\,dx,$$

где $\rho(x)\geqslant 0$—заданная *весовая функция*, принимающая бóльшие значения для тех x, которые должны сыграть бóльшую роль.

Функции $g(x)$ и $h(x)$ называются *ортогональными с весом* ρ на интервале (a, b), если

$$\int_a^b f(x)\,g(x)\,\rho(x)\,dx=0.$$

Вся общая теория рядов по ортогональным функциям (п. 20, 21, 26, 27) непосредственно переносится на функции, ортогональные с весом; при этом во всех участвующих интегралах надо под знаком интеграла дописать весовой множитель $\rho(x)$.

Интеграл с весом ρ легко преобразовать в интеграл без веса (т. е. с весом 1) при помощи замены независимой переменной

$$\rho(x)\,dx=d\overline{x},\quad \overline{x}=\int_{x_0}^{x}\rho(x)\,dx=\overline{x}(x),\quad \overline{x}(a)=\overline{a},\quad \overline{x}(b)=\overline{b}, \tag{126}$$

так как если обозначить $f(x)=f(x(\overline{x}))=\overline{f}(\overline{x})$, то получим

$$\int_a^b f(x)\,\rho(x)\,dx=\int_{\overline{a}}^{\overline{b}}\overline{f}(\overline{x})\,d\overline{x}.$$

При таком преобразовании любая система функций, ортогональных с весом ρ, переходит в систему функций, ортогональных в обычном смысле (п. 20). Тем не менее иногда удобнее такой замены не делать, а рассматривать все интегрирования с весом.

Обратно, с помощью замены (126) можно из любой ортогональной системы функций получить систему, ортогональную с весом ρ. Так как при таком преобразовании разложения в ряды также соответственно преобразуются, то полная система после этой замены остается полной. Например, исходя из полной ортогональной системы функций

$$1,\ \cos\bar{x},\ \cos 2\bar{x},\ \ldots,\ \cos n\bar{x},\ \ldots \qquad (0 \leqslant \bar{x} \leqslant \pi)$$

(см. систему (94)) и совершая преобразование

$$\bar{x} = \pi - \arccos x\ (-1 \leqslant x \leqslant 1), \qquad d\bar{x} = \frac{1}{\sqrt{1-x^2}}\,dx,$$

получим после несущественной перемены знаков (проверьте!)

$$T_0(x) = 1,\ T_1(x) = \cos\arccos x = x,\ T_2(x) = \cos(2\arccos x) = 2x^2 - 1,$$
$$T_3(x) = \cos(3\arccos x) = 4x^3 - 3x,\ \ldots,\ T_n(x) = \cos(n\arccos x),\ \ldots$$

Эти многочлены были введены Чебышевым в 1857 г., они и называются *многочленами Чебышева*. Мы видим, что они образуют полную систему функций, ортогональных с весом $\frac{1}{\sqrt{1-x^2}}$ на отрезке $-1 \leqslant x \leqslant 1$. Их можно получить также из системы функций (97) с помощью процесса ортогонализации с данным весом (проделайте это!).

Избавиться от весовой функции можно также с помощью следующего простого замечания: если система функций

$$g_1(x),\ g_2(x),\ \ldots,\ g_n(x),\ \ldots \tag{127}$$

ортогональна на интервале $a < x < b$ с весом $\rho(x)$, то система функций

$$g_1(x)\sqrt{\rho(x)},\quad g_2(x)\sqrt{\rho(x)},\quad \ldots,\quad g_n(x)\sqrt{\rho(x)},\ \ldots \tag{128}$$

ортогональна на том же интервале без веса (проверьте!); чтобы разложить функцию $f(x)$ в ряд по системе (127), достаточно разложить функцию $f(x)\sqrt{\rho(x)}$ в ряд по системе (128), а затем сократить результат на $\sqrt{\rho(x)}$.

30. Кратные ряды Фурье. При разложении в ряды функций от нескольких аргументов обычно применяются системы функций, зависящих от нескольких индексов, число которых равно числу аргументов. Тогда ряды получаются кратными, как в п. 17.

Теория кратных рядов по ортогональным функциям развивается совершенно аналогично теории простых рядов. Рассмотрим для простоты случай двух аргументов. Тогда ортогональные в некоторой области D функции, образующие полную систему, должны зависеть от двух индексов, т. е. иметь вид $\varphi_{mn}(x, y)$, где m и n принимают некоторые дискретные значения. Ряд по такой системе функций имеет вид

$$f(x, y) = \sum_{m,\,n} a_{mn}\varphi_{mn}(x, y);$$

здесь подразумевается двойная сумма. Коэффициенты вычисляются, как в п. 21:

$$a_{mn} = \frac{\displaystyle\iint_D f(x, y)\,\varphi_{m,n}(x, y)\,dx\,dy}{\displaystyle\iint_D \varphi_{m,n}^2(x, y)\,dx\,dy}. \tag{129}$$

Обычно примеры систем ортогональных функций от нескольких аргументов строятся по следующему принципу. Пусть даны две полные ортогональные системы функций одного переменного

$$g_1(x),\ g_2(x),\ \ldots,\ g_n(x),\ \ldots\ (a \leqslant x \leqslant b) \text{ и } h_1(x),\ h_2(x),\ \ldots,\ h_n(x),\ \ldots \quad (c \leqslant x \leqslant d).$$

Тогда система функций

$$\varphi_{mn}(x,y) = g_m(x)\,h_n(y) \qquad (m,\ n = 1,\ 2,\ \ldots) \tag{130}$$

будет ортогональной и полной в прямоугольнике П: $a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$.

Ортогональность этой системы вытекает из того, что

$$\iint_{\Pi} \varphi_{mn}(x,\ y)\,\varphi_{\overline{m}\,\overline{n}}(x,\ y)\,dx\,dy = \int_a^b dx \int_c^d g_m(x)\,h_n(y)\,g_{\overline{m}}(x)\,h_{\overline{n}}(y)\,dy =$$

$$= \int_a^b g_m(x)\,g_{\overline{m}}(x)\,dx \int_c^d h_n(y)\,h_{\overline{n}}(y)\,dy,$$

т. е. получается нуль, за исключением случая, когда $m = \overline{m}$ и одновременно $n = \overline{n}$. Полнота следует из того, что произвольную функцию $f(x,\ y)$, заданную в П, можно при любом фиксированном x разложить по функциям $h_n(y)$:

$$f(x,\ y) = \sum_{n=1}^{\infty} A_n(x)\,h_n(y),$$

где коэффициенты разложения зависят от x. Если эти коэффициенты $A_n(x)$ разложить по функциям $g_m(x)$, то и получится разложение функции $f(x,\ y)$ в двойной ряд по функциям (130):

$$f(x,\ y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} g_m(x)\,h_n(y),$$

которое, таким образом, возможно.

Так, исходя из растянутой системы функций (95), можно построить такую полную ортогональную систему функций на прямоугольнике $0 \leqslant x \leqslant l_1$, $0 \leqslant y \leqslant l_2$:

$$\varphi_{mn}(x,\ y) = \sin\frac{m\pi x}{l_2}\sin\frac{n\pi y}{l_2} \qquad (m,\ n = 1,\ 2,\ 3,\ \ldots).$$

Разложение по этой системе имеет вид

$$f(x,\ y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} \sin\frac{m\pi x}{l_1}\sin\frac{n\pi y}{l_2} \qquad (0 \leqslant x \leqslant l_1,\ 0 \leqslant y \leqslant l_2).$$

Коэффициенты этого ряда в силу формулы (129) равны (проверьте!)

$$a_{mn} = \frac{4}{l_1 l_2} \int_0^{l_1} dx \int_0^{l_2} f(x,\ y) \sin\frac{m\pi x}{l_1}\sin\frac{n\pi y}{l_2}\,dy.$$

31. Приложение к уравнению колебаний конечной струны. Ряды Фурье имеют многочисленные приложения к решению уравнений математической физики. Мы здесь покажем только один пример такого приложения к решению уравнения малых свободных поперечных колебаний натянутой струны.

Будем пренебрегать весом струны и направим ось x вдоль положения равновесия струны так, чтобы концы струны имели абсциссы $x=0$ и $x=l$; будем считать колебания происходящими в плоскости и обозначим через $u=u(x, t)$ поперечное отклонение точки с абсциссой x в момент t. Тогда в курсах уравнений математической физики доказывается, что функция $u(x, t)$ должна удовлетворять уравнению с частными производными

$$\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2}; \tag{131}$$

здесь a — постоянная, $a=\sqrt{\frac{T}{\rho}}$, где T — сила натяжения струны, а ρ — ее линейная плотность. Условие закрепления концов струны можно записать в виде

$$u\,|_{x=0}=0, \quad u\,|_{x=l}=0 \qquad (\text{при всех } t); \tag{132}$$

кроме того, будем считать, что известны отклонения и скорости точек струны в начальный момент времени:

$$u\Big|_{t=0}=\varphi(x), \quad \frac{\partial u}{\partial t}\Big|_{t=0}=\psi(x). \tag{133}$$

Таким образом, получается следующая математическая задача: нужно решить уравнение (131) при *граничных условиях* (132) и *начальных условиях* (133).

Для решения этой задачи разложим искомое решение при каждом фиксированном $t\geqslant 0$ в ряд вида (107); когда коэффициенты будут зависеть от t, т. е. мы получаем

$$u(x, t)=\sum_{n=1}^{\infty} b_n(t)\sin\frac{\pi n x}{l}. \tag{134}$$

Чтобы найти коэффициенты $b_n(t)$, подставим это выражение в уравнение (131). Мы получим

$$\sum_{n=1}^{\infty} b_n''(t)\sin\frac{\pi n x}{l}=-a^2\sum_{n=1}^{\infty} b_n(t)\frac{\pi^2 n^2}{l^2}\sin\frac{\pi n x}{l},$$

т. е.

$$b_n''(t)=-\frac{a^2\pi^2 n^2}{l^2}b_n(t).$$

Решив это обыкновенное линейное уравнение с постоянными коэффициентами по методу п. XV.17, получим

$$b_n(t)=A_n\cos\frac{a\pi n}{l}t+B_n\sin\frac{a\pi n}{l}t,$$

т. е. в силу (134)

$$u=\sum_{n=1}^{\infty}\left(A_n\cos\frac{a\pi n}{l}t+B_n\sin\frac{a\pi n}{l}t\right)\sin\frac{\pi n}{l}x. \tag{135}$$

Чтобы найти коэффициенты A_n, B_n, воспользуемся начальными условиями (133); получим (проверьте!)

$$\varphi(t)=\sum_{n=1}^{\infty}A_n\sin\frac{\pi n}{l}x, \quad \psi(t)=\sum_{n=1}^{\infty}B_n\frac{a\pi n}{l}\sin\frac{\pi n}{l}x \qquad (0\leqslant x\leqslant l).$$

Получилось опять разложение в ряд Фурье вида (107), из которого находим

$$A_n=\frac{2}{l}\int_0^l \varphi(t)\sin\frac{\pi n x}{l}\,dx, \qquad B_n=\frac{2}{a\pi n}\int_0^l \psi(t)\sin\frac{\pi n x}{l}\,dx.$$

Подстановка этих величин в (135) и даст искомое решение. При этом граничные условия (132) удовлетворяются автоматически.

§ 5. Преобразование Фурье

32. Формулы преобразования Фурье. Будем исходить из формулы (115), которая дает представление любой конечной функции $f(x)$ на интервале $-l<x<l$. В этом представлении участвуют *гармоники*, т. е. функции вида e^{ikx} с *волновыми числами* k, равными

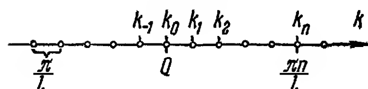


Рис. 342.

$$k=k_n=\frac{\pi n}{l} \tag{136}$$

$(n=\ldots,\ -2,\ -1,\ 0,\ 1,\ 2,\ \ldots)$.

Набор этих значений, изображенный на рис. 342, называется *спектром* волновых чисел; он *дискретен*, т. е. состоит из отдельных точек, каждой из которых в разложении (115) отвечает гармоника e^{ik_nx} с *комплексной амплитудой* c_n, определяемой по формуле (116):

$$c_n=\frac{1}{2l}\int_{-l}^{l} f(x)\,e^{-ik_nx}dx. \tag{137}$$

Допустим сначала для простоты, что функция $f(x)$ вне некоторого *конечного* интервала $a\leqslant x\leqslant b$ тождественно равна нулю, и обозначим

$$\hat{f}(k)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\,e^{-ikx}dx; \tag{138}$$

фактически этот интеграл берется лишь по интервалу $a\leqslant x\leqslant b$. Тогда при достаточно больших l формулу (137) для коэффициентов Фурье можно переписать в виде

$$c_n=\frac{\pi}{l}\,\frac{1}{2\pi}\int_a^b f(x)\,e^{-ik_nx}dx=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\,e^{-ik_nx}dx\,\frac{\pi}{l}=\hat{f}(k_n)\,\Delta k, \tag{139}$$

где $\Delta k=\dfrac{\pi}{l}$ — расстояние между соседними волновыми числами в спектре (см. формулу (136) или рис. 342). Поэтому формулу (115) для представления $f(x)$ можно переписать в виде

$$f(x)=\sum_n c_n e^{ik_nx}=\sum_n \hat{f}(k_n)\,e^{ik_nx}\Delta k \qquad (-l<x<l). \tag{140}$$

Допустим теперь, что l весьма велико. Тогда спектр (рис. 342) становится очень густым, т. е. Δk — очень малым. В пределе при $l \to \infty$ сумма (140), которая является интегральной суммой, переходит в интеграл, т. е. мы получаем

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(k)\, e^{ikx}\, dk \qquad (-\infty < x < \infty). \tag{141}$$

В этом представлении волновое число k принимает *все* значения от $-\infty$ до ∞, т. е. *в пределе при $l \to \infty$ дискретный спектр волновых чисел переходит в непрерывный спектр.*

Из первого равенства (139) видно, что при $l \to \infty$ амплитуды $c_n \to 0$, т. е. в пределе каждая отдельно взятая гармоника присутствует с нулевой амплитудой. Формула (141) означает, что в пределе амплитуда оказывается «размазанной» по всему непрерывному спектру волновых чисел, так что на каждый бесконечно малый интервал от k до $k+dk$ приходится бесконечно малая амплитуда

$$dc = \hat{f}(k)\, dk. \tag{142}$$

Ситуация здесь несколько напоминает ту, которая получается при переходе от дискретной модели материального тела к его непрерывной модели, когда принимается, что масса каждой отдельно взятой точки равна нулю, а масса оказывается «размазанной» по всем точкам с определенной плотностью. Подобно этому формула (142) означает, что $\hat{f}(k)$ — это плотность амплитуды гармоник на бесконечно малом интервале волновых чисел, причем плотность берется в расчете на единицу длины этого интервала; поэтому функция $\hat{f}(k)$ называется также *спектральной плотностью* функции $f(x)$.

Формулы (138) и (141) называются формулами *преобразования Фурье*, причем (138) — *прямой* формулой, а (141) — *обратной*. При их выводе мы предполагали, что функция f вне некоторого конечного интервала тождественно равна нулю; такие функции называются *финитными*. Более детальное исследование показывает, что формулы справедливы и в том случае, если интеграл (138) понимается как несобственный, но для его сходимости надо потребовать, чтобы

$$\int_{-\infty}^{\infty} |f(x)|\, dx < \infty. \tag{143}$$

Таким образом, *каждой функции* $f(x)$, *удовлетворяющей условию* (143), *отвечает ее «фурье-образ»* $\hat{f}(k)$ (т. е. результат преобразования Фурье), *который определяется формулой* (138); *обратно, функция* $f(x)$ *выражается через свой фурье-образ по формуле* (141).

Если функция $f(x)$ четная, то при вычислении интеграла (138) можно воспользоваться свойством 9 п. XIV.4, и мы получаем

$$\hat{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \cos kx\, dx - \frac{i}{2\pi} \int_{-\infty}^{\infty} f(x) \sin kx\, dx = \frac{1}{\pi} \int_{0}^{\infty} f(x) \cos kx\, dx.$$

Отсюда функция $\hat{f}(k)$ также четная, и из (141) мы получаем

$$f(x) = 2 \int_{0}^{\infty} \hat{f}(k) \cos kx\, dk;$$

это — формулы *косинус-преобразования Фурье. Если функция $f(x)$ нечетная,* то мы подобным образом получаем формулы *синус-преобразования Фурье*:

$$i\hat{f}(k)=\frac{1}{\pi}\int_0^\infty f(x)\sin kx\,dx,\quad f(x)=2\int_0^\infty i\hat{f}(k)\sin kx\,dk;$$

впрочем, в этом случае фурье-образом от $f(x)$ обычно называется не $\hat{f}(k)$, а $i\hat{f}(k)$.

Если функция $f(x)$ рассматривается на интервале $0<x<\infty$, то ее можно продолжить на интервал $-\infty<x<0$ как четным, так и нечетным образом. Поэтому если считать как x, так и k положительным, то можно пользоваться как косинус-, так и синус-преобразованиями Фурье; однако образы, полученные при этих преобразованиях, будут, вообще говоря, различными.

Рассмотрим пример. Пусть $f(x)$ — четная функция, равная 1 на интервале $-1<x<1$ и нулю вне его. По формуле косинус-преобразования Фурье

$$\hat{f}(k)=\frac{1}{\pi}\left(\int_0^1 1\cdot\cos kx\,dx+\int_1^\infty 0\cdot\cos kx\,dx\right)=\frac{\sin k}{\pi k}.$$

Применяя обратную формулу, получаем

$$f(x)=2\int_0^\infty\frac{\sin k}{\pi k}\cos kx\,dk. \tag{144}$$

Подобно ряду Фурье (п. 25), если подставлять в формулу для обратного преобразования числовые значения x, мы получим значение $f(x)$ во всех точках непрерывности этой функции и значение $[f(x-0)+f(x+0)]/2$ во всех точках, где она имеет конечный скачок. В частности, подставляя в (144) значение $x=0$, при котором функция f непрерывна, мы получим

$$1=2\int_0^\infty\frac{\sin k}{\pi k}\,dk,\quad\text{откуда}\quad\int_0^\infty\frac{\sin k}{k}\,dk=\frac{\pi}{2}.$$

Интегральная формула, равносильная формулам преобразования Фурье, была получена Фурье в 1811 г.

33. Свойства преобразования Фурье. Преобразование Фурье обладает рядом полезных свойств; некоторые из них мы здесь перечислим. Прежде всего, ясно, что преобразование Фурье является *оператором* (п. XIV. 26), для которого функция $f(x)$ является прообразом, а функция $\hat{f}(k)$ — образом.

1. *Оператор Фурье является линейным,* т. е.

$$\widehat{(f_1+f_2)}=\hat{f}_1+\hat{f}_2,\quad\widehat{\alpha f}=\alpha\hat{f}\qquad(\alpha=\text{const}). \tag{145}$$

Это сразу следует из формулы (138) и из того, что интегрирование является линейной операцией.

Из формулы (145) вытекает, в частности, что если функция f зависит не только от x, но и от некоторого параметра t и потому $\hat{f}$ также зависит от этого параметра, то

$$\widehat{\frac{f_{t+\Delta t}-f_t}{\Delta t}}=\frac{\hat{f}_{t+\Delta t}-\hat{f}_t}{\Delta t}.$$

Переходя к пределу при $\Delta t \to 0$, получаем, что

$$\widehat{\frac{\partial f}{\partial t}} = \frac{\partial \widehat{f}}{\partial t},$$

т. е. *производная по параметру от прообраза преобразуется в производную по параметру от образа*. Доказательство показывает, что *это свойство справедливо для любого линейного оператора*.

2. *Если функцию f продифференцировать по x, то ее образ $\widehat{f}$ умножится на ik*. В самом деле, образом функции $f'(x)$ служит

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} f'(x)\, e^{-ikx}\, dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\, df(x) =$$

$$= \frac{1}{2\pi} e^{-ikx} f(x)\Big|_{x=-\infty}^{\infty} - \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)(-ik)\, e^{-ikx}\, dx$$

(мы произвели интегрирование по частям). Однако из условия (143) следует, что $f(\pm\infty) = 0$, и потому первое слагаемое в правой части пропадает. Но второе равно

$$\frac{ik}{2\pi}\int_{-\infty}^{\infty} f(x)\, e^{-ikx}\, dx = ik\widehat{f}(k),$$

что и требовалось доказать.

Преобразуя подобным образом обратную формулу (141), можно доказать, что *если функцию $\widehat{f}$ продифференцировать, то ее прообраз f умножится на $-ik$*.

3. *Если функция $f(x)$ преобразуется в $\widehat{f}(k)$, то функция $f(\alpha x)$* ($\alpha = \text{const} > 0$) *преобразуется в* $\frac{1}{\alpha}\widehat{f}\left(\frac{k}{\alpha}\right)$. В самом деле,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} f(\alpha x)\, e^{-ikx}\, dx = |\,\alpha x = s\,| = \frac{1}{\alpha}\,\frac{1}{2\pi}\int_{-\infty}^{\infty} f(s)\, e^{-i\frac{k}{\alpha}s}\, ds = \frac{1}{\alpha}\widehat{f}\left(\frac{k}{\alpha}\right).$$

Таким образом, при растяжении прообраза в несколько раз вдоль оси независимой переменной образ сжимается во столько же раз. Это значит, что невозможно одновременно *локализовать*, т. е. сосредоточить на оси независимой переменной, как функцию-прообраз, так и ее спектральную плотность. Этот *принцип неопределенности* имеет многочисленные применения в физике.

4. *Если функцию $f(x)$ сдвинуть на* $\beta = \text{const}$, *то ее образ умножится на* $e^{-i\beta k}$. В самом деле,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x-\beta)\, e^{-ikx}\, dx = |\,x-\beta = s\,| = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(s)\, e^{-iks} e^{-ik\beta}\, ds = e^{-ik\beta}\widehat{f}(k).$$

Обратно, *если образ сдвинуть на β, то прообраз умножится на* $e^{i\beta x}$.

5. *Равенство Парсеваля*. Если формулу (123) применить к ряду (115), мы получим

$$\int_{-l}^{l} |f(x)|^2\, dx = 2l \sum_{n=-\infty}^{\infty} |c_n|^2.$$

Пользуясь формулой (139), получаем отсюда

$$\int_{-l}^{l} |f(x)|^2 dx = 2l \sum_{n=-\infty}^{\infty} |\hat{f}(k_n)|^2 (\Delta k)^2 = 2\pi \sum_n |\hat{f}(k_n)|^2 \Delta k.$$

Переходя к пределу при $l \to \infty$, мы подобно п. 32 получаем

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = 2\pi \int_{-\infty}^{\infty} |\hat{f}(k)|^2 dk.$$

Это и есть *равенство Парсеваля для преобразования Фурье.*

34. Приложение к уравнению колебаний бесконечной струны. Интегральное преобразование Фурье применяется к решению задач математической физики для бесконечных сред. Покажем это на примере уравнения (131), но теперь будем струну считать бесконечной, т. е. $-\infty < x < \infty$, так что граничные условия отпадают; в начальных условиях (133) мы для простоты будем считать $\psi(x) \equiv 0$. Обозначим через $\hat{u}(k, t)$ фурье-образ решения при любом фиксированном $t \geqslant 0$. Совершая преобразование по Фурье обеих частей уравнения (131), на основе свойств 1 и 2 п. 33 получим

$$\frac{\partial^2 \hat{u}}{\partial t^2} = a^2 (ik)^2 \hat{u} = -a^2 k^2 \hat{u}.$$

При любом фиксированном k—это обыкновенное линейное дифференциальное уравнение с постоянными коэффициентами, откуда стандартным методом п. XV.17 получаем

$$\hat{u} = C_1(k) e^{-iakt} + C_2(k) e^{iakt}. \tag{146}$$

Теперь преобразуем по Фурье начальные условия (133):

$$\hat{u}\Big|_{t=0} = \hat{\varphi}(k), \qquad \frac{\partial \hat{u}}{\partial t}\Big|_{t=0} = 0.$$

Поэтому формулы (146) дают (проверьте!)

$$\hat{u} = \frac{1}{2} \hat{\varphi}(k) e^{-iatk} + \frac{1}{2} \hat{\varphi}(k) e^{iatk}.$$

На основе свойств 1 и 4 п. 33 возвращаемся к прообразу:

$$u = \frac{1}{2} \varphi(x - at) + \frac{1}{2} \varphi(x + at).$$

Это и есть искомое решение. Оно имеет весьма простой смысл: начальное отклонение делится пополам и одна половина в момент t сдвигается на расстояние at в положительном направлении оси x, а другая—на то же расстояние в отрицательном направлении. Другими словами, по струне бегут две волны со скоростью a, в положительном и в отрицательном направлениях, без изменения своей формы, причем наблюдается все время результат сложения этих волн. Тем самым выяснен физический смысл постоянной a в уравнении (131): она равна скорости распространения возмущений по струне.

ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. Случайные события и их вероятности

1. Случайные события. *Теория вероятностей — это наука о случайных событиях.* Понятие события принадлежит к числу основных и потому с трудом поддающихся определению. На первом этапе *событием можно считать все то, что может произойти или не произойти при осуществлении определенного комплекса условий*; каждое такое осуществление называется *испытанием*. Например, событие может состоять в выпадании герба при бросании монеты; в этом случае испытанием будет служить бросание монеты. Событие может состоять в том, что некоторое изделие, выбранное из партии готовых изделий, окажется бракованным; в этом случае испытанием будет служить акт *выборки* изделия из партии. Однако не обязательно «испытание» в том смысле, как это понимается здесь, надо связывать с человеческой деятельностью. Например, событие может состоять в том, что в какой-либо день над Байкалом пройдет дождь; тогда испытание будет состоять просто в наступлении дня.

Характерной чертой *случайного* события является то, что в результате испытания оно происходит не обязательно; это отличает случайное событие от *детерминированного*, которое происходит обязательно. Случайность события связана с тем, что многие факторы, сопутствующие испытанию и существенные для его исхода, не задаются. Эта неполнота информации в одних случаях является принципиальной (например, в азартных играх или в военных действиях) или недоступной современному уровню развития науки (например, при прогнозировании погоды). Предположение о принципиальной непредсказуемости результатов отдельных испытаний лежит в основе ряда наук, таких как квантовая механика, генетика, социология и т. д. В других случаях точное предсказание результата испытаний является принципиально возможным, но нецелесообразным практически, так как оно потребовало бы неоправданных затрат на дополнительные точные и сверхточные измерения и т. п.

Закономерности случайных событий проявляются при многократном повторении испытаний. Например, нельзя предсказать результат единичного бросания монеты: может выпасть как герб, так и цифра. Никого особенно не удивит, если при десятикратном бросании герб выпадет всего два раза. Но если при 1000-кратном бросании герб

выпадает всего 200 раз, то всякий с основанием скажет, что что-то не в порядке либо с монетой, либо с бросанием: ведь при симметричных условиях ни герб, ни цифра не имеют преимуществ друг перед другом, т. е. они должны выпадать примерно одинаково часто. Конечно, после 1000 бросаний герб не обязательно выпадает ровно 500 раз, он может выпасть и 490, и 525 раз, но не 200! Подобным образом результат однократной выборки из партии изделий не позволяет сделать заключение о качестве партии, это можно сделать только при многократной выборке, как говорят, при большом *объеме выборки*, и т. п. Поэтому, уточняя первую фразу этого пункта, часто говорят не вообще о случайных событиях, а о *массовых* случайных событиях, понимая под массовостью многократную повторяемость.

Отметим, что повторные испытания можно понимать двояко. Например, можно 1000 раз подбросить одну и ту же монету, но можно и разновременно или даже одновременно подбросить независимо друг от друга 1000 одинаковых монет, — это совершенно равноценно. Мы не будем делать различия между этими видами повторений.

2. Вероятность. Даже в обыденной жизни мы часто называем одно событие очень вероятным, а другое маловероятным; при возможности многократного повторения испытаний это означает, конечно, что первое событие будет происходить часто, а второе — редко. Важнейшей чертой теории вероятностей является то, что в ней говорится не просто о большой или малой вероятности события, а о точном численном значении этой вероятности, т. е. *вероятность считается величиной*, характеризующей частоту наступления события при многократном повторении испытаний.

Допустим, что монету бросали 1000 раз и герб выпал 490 раз. Тогда отношение $\frac{490}{1000}=0{,}49$ называется *относительной частотой* выпадания герба в данной серии испытаний. Пусть монету бросали 10 000 раз и герб выпал 5027 раз; тогда относительная частота будет равна 0,5027. Ясно, что если монета симметричная и число бросаний увеличивается, то относительная частота выпаданий герба должна приближаться к 0,5, так как ни одна сторона монеты не должна иметь предпочтения перед другой. Это число 0,5 и называется *вероятностью* выпадания герба при бросании монеты.

В общем случае определение аналогичное. Пусть случайное событие, которое мы обозначим буквой A, после осуществления серии из N независимых испытаний, произошло N_A раз; тогда отношение $\frac{N_A}{N}$ называется *относительной частотой события A в данной серии испытаний. Предел*

$$\mathbf{P}\{A\}=\lim_{N\to\infty}\frac{N_A}{N},$$

к которому стремится относительная частота события А при неограниченном увеличении числа испытаний, называется вероятностью случайного события А. Буква **P** происходит от французского probabilité — вероятность.

Таким образом, если число испытаний велико, то относительную частоту события практически можно просто принять за его вероятность. Это дает метод *эмпирического подсчета вероятностей* в тех случаях, когда их теоретический подсчет затруднен. Например, из статистики рождений известна с большой точностью вероятность рождения мальчика: она равна 0,512, хотя в отдельные годы и несколько отклоняется от этой величины. Это число не дает возможности в каждом отдельном случае предсказать, кто именно родится — мальчик или девочка, можно только сказать, что появление мальчика чуть вероятней. Но можно утверждать, что из миллиона родившихся детей число мальчиков будет близко к 512 тысячам (насколько именно близко, можно заключить после прочтения п. 20).

В некоторых случаях вероятность можно вычислить с помощью *подсчета шансов*. Поясним этот метод на примере. Пусть бросается игральный кубик, грани которого отвечают числу очков от 1 до 6, и надо подсчитать вероятность того, что выпавшее число очков будет делиться на 3. Произведем большое число N бросаний и обозначим через N_k ($k=1, 2, 3, 4, 5, 6$) число появлений k очков. Тогда $N_1+N_2+\ldots+N_6=N$, т. е. $\frac{N_1}{N}+\frac{N_2}{N}+\ldots+\frac{N_6}{N}=1$. Но так как ни одно число очков при своем появлении не имеет предпочтения перед другим, то все эти шесть дробей приближенно, а в пределе точно равны друг другу, т. е. равны $\frac{1}{6}$. Однако выпавшее число очков будет делиться на три, если оно равно трем или шести, т. е. в N_3+N_6 случаях; относительная частота этого события равна

$$\frac{N_3+N_6}{N}=\frac{N_3}{N}+\frac{N_6}{N}\underset{N\to\infty}{\longrightarrow}2\cdot\frac{1}{6}=\frac{1}{3}.$$

Это и есть искомая вероятность. Полученный результат выражают короче такими словами: при бросании кубика имеется шесть шансов в соответствии с числом выпавших очков, из них два *благоприятных шанса*, отвечающих трем и шести очкам, остальные *неблагоприятные*; значит, вероятность рассматриваемого события равна $\frac{2}{6}=\frac{1}{3}$.

Общая схема такого подсчета такова. Пусть в результате испытания должен произойти какой-либо *один и только один* из n исходов, результатов, *причем все эти исходы равновозможны*. Тогда эти исходы называются *шансами*. Пусть рассматриваемое событие A происходит при каких-либо q исходах, которые называются *благоприятными шансами*, и не происходит при остальных $n-q$ *неблаго-*

приятных шансах. Тогда рассуждение, аналогичное проведенному в предыдущем абзаце, показывает, что

$$\mathbf{P}\{A\}=\frac{q}{n},$$

т. е. *вероятность события равна отношению числа благоприятных шансов к числу всех шансов.*

С помощью шансов легко подсчитать вероятность выигрыша в лотерее (надо разделить число выигрышей на число всех билетов) и другие подобные вероятности. При этом надо тщательно следить за тем, чтобы шансы были действительно равновозможными, т. е. равновероятными. Например, нельзя было рассуждать так: число очков, выпавшее при бросании кубика, может либо делиться на три, либо нет, а потому для первого случая имеется один шанс из двух, т. е. его вероятность равна $\frac{1}{2}$ (в чем здесь ошибка?). Конечно, предположение о равновозможности шансов является схематизацией действительности и в реальных случаях выполняется лишь с большей или меньшей степенью точности. Оно возможно лишь при определенной *симметрии* рассматриваемого испытания.

Схема с шансами имеет следующий вариант, который мы поясним на примере. Пусть на поверхности однородного шарика зачернена некоторая фигура, причем площадь всей поверхности равна S, а площадь фигуры — S_0. Пусть этот шарик произвольно бросается на горизонтальную плоскость и требуется подсчитать вероятность того, что шарик ударится о плоскость зачерненной частью своей поверхности. Для подсчета вероятности допустим, что вся поверхность разбита на маленькие участки равной площади dS. Тогда точка столкновения может принадлежать любому из этих участков, причем с равной вероятностью. Однако всего имеется $\frac{S}{dS}$ участков, из них $\frac{S_0}{dS}$ зачернено. Значит, для удара зачерненной частью имеется $\frac{S_0}{dS}$ шансов из $\frac{S}{dS}$, а потому искомая вероятность равна $\frac{S_0}{dS}:\frac{S}{dS}=\frac{S_0}{S}$. Интересно, что решение аналогичной задачи с эллипсоидом вместо шара зависит не только от площади, но и от расположения зачерненной фигуры и получается с помощью интегрирования, что мы предоставляем сделать читателю.

В заключение отметим, что в обиходе термин «вероятность» иногда употребляется по отношению к событиям, которые даже мысленно нельзя считать повторяющимися: например, говорят о вероятности жизни на Марсе. Здесь лучше говорить не о вероятности, а о *степени правдоподобия* гипотезы. Теория правдоподобия разработана пока еще недостаточно.

3. Основные свойства вероятности. 1. *Вероятность любого события есть безразмерная величина, заключенная между* 0 *и* 1:

$$0\leqslant\mathbf{P}\{A\}\leqslant 1.$$

Это сразу вытекает из определения вероятности (п. 2), из которого также видно, что чем $\mathbf{P}\{A\}$ больше, тем событие возможней, вероятней в обычном понимании этого термина.

2. *Вероятность достоверного события*, т. е. такого, которое при испытании обязательно произойдет, *равна единице.* Таким образом, мы считаем достоверное, детерминированное событие частным случаем случайного, наподобие того как в п. I.5 мы объявили постоянную величину частным случаем переменной. Аналогично *вероятность невозможного события равна нулю.*

Если имеется лишь конечное число шансов, то справедливы и обратные утверждения: если вероятность равна единице (нулю), то событие достоверно (невозможно). Однако в общем случае здесь имеется некоторое различие. Например, из п. 2 вытекает, что вероятность падения произвольно брошенного шарика на *заранее заданную* точку плоскости равна нулю. Хотя такое событие и не противоречит законам механики, но практически является совершенно невозможным.

3. *Сумма вероятностей двух противоположных событий равна единице*. При этом два события называются *противоположными*, если наступление одного из них равносильно ненаступлению другого, иначе говоря, каждое из этих событий является отрицанием другого. Например, если вероятность попадания в мишень при каких-либо условиях равна 0,2, то вероятность промаха при тех же условиях равна 0,8. Для доказательства этого утверждения в общем случае обозначим через A и $\overline{A}$ два взаимно противоположных события. Пусть произведено N испытаний и при этом N_A раз произошло событие A и $N_{\overline{A}}$ раз — событие $\overline{A}$. Тогда ясно, что $N_A+N_{\overline{A}}=N$, откуда $\frac{N_A}{N}+\frac{N_{\overline{A}}}{N}=1$. Переходя к пределу при $N\to\infty$, получаем, что

$$\mathbf{P}\{A\}+\mathbf{P}\{\overline{A}\}=1. \tag{1}$$

4. Аналогично доказывается более общее утверждение: *если в результате испытания должно произойти одно и только одно из некоторых событий* $A_1, A_2, \ldots, A_k$, то

$$\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}+\ldots+\mathbf{P}\{A_k\}=1. \tag{2}$$

5. Пусть каждое из двух событий A, B может произойти или не произойти в результате одного и того же испытания. Пусть произведено N испытаний; обозначим через $N_{A \text{ и } B}$ число испытаний, при которых оба события произошли, через $N_{A \text{ и } \overline{B}}$ — число испытаний, при которых A произошло, а B не произошло, и т. д. Тогда

$$N=N_{A \text{ и } B}+N_{A \text{ и } \overline{B}}+N_{\overline{A} \text{ и } B}+N_{\overline{A} \text{ и } \overline{B}}.$$

Кроме того, общее число испытаний, при которых событие A произошло, равно

$$N_A=N_{A \text{ и } B}+N_{A \text{ и } \overline{B}}, \quad \text{аналогично} \quad N_B=N_{A \text{ и } B}+N_{\overline{A} \text{ и } B}. \tag{3}$$

Далее, обозначим через $N_{A \text{ или } B}$ число испытаний, при которых произошло *по крайней мере* одно из событий A и B; тогда

$$N_{A \text{ или } B} = N_{A \text{ и } B} + N_{A \text{ и } \bar{B}} + N_{\bar{A} \text{ и } B}. \tag{4}$$

Из формул (3)—(4) получаем

$$\frac{N_{A \text{ или } B}}{N} = \frac{N_A}{N} + \frac{N_B}{N} - \frac{N_{A \text{ и } B}}{N}.$$

Переходя к пределу при $N \to \infty$, получаем формулу

$$\mathbf{P}\{A \text{ или } B\} = \mathbf{P}\{A\} + \mathbf{P}\{B\} - \mathbf{P}\{A \text{ и } B\},$$

в которой смысл отдельных членов естественно расшифровывается.

В частности, если события A и B *несовместны*, т. е. не могут произойти одновременно, получаем *правило сложения вероятностей:*

$$\mathbf{P}\{A \text{ или } B\} = \mathbf{P}\{A\} + \mathbf{P}\{B\} \qquad (A \text{ и } B \text{ несовместны}).$$

Аналогично доказывается более общее правило: если события A_1, A_2, ..., A_k *попарно несовместны*, то

$$\mathbf{P}\{A_1, \text{ или } A_2, \ldots, \text{ или } A_k\} = \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\} + \ldots + \mathbf{P}\{A_k\}. \tag{5}$$

4. Правило умножения вероятностей. Пусть A и B — два события. Тогда *условной вероятностью* $\mathbf{P}\{A|B\}$ события A при условии B называется вероятность события A, вычисленная в предположении, что событие B произошло. Таким образом, при подсчете этой вероятности с помощью относительной частоты (п. 2) отбираются только те испытания, при которых событие B произошло:

$$\mathbf{P}\{A|B\} = \lim_{N \to \infty} \frac{N_{A \text{ и } B}}{N_B}.$$

Пусть, например, имеется две урны, причем в первой содержатся три черных шара и один белый, а во второй — один черный и три белых. Мы наугад подходим к одной из урн и вынимаем один шар; какова вероятность, что он будет черным? Из соображений симметрии ясно, что $\mathbf{P}\{A_ч\} = \frac{1}{2}$, где $A_ч$ — событие, состоящее в появлении черного шара. Но пусть известно, что мы подошли именно к первой урне; обозначим это событие через B_1; тогда ясно, что соответствующая условная вероятность $\mathbf{P}\{A_ч|B_1\} = \frac{3}{4}$.

Из простой формулы

$$\frac{N_{A \text{ и } B}}{N} = \frac{N_B}{N} \frac{N_{A \text{ и } B}}{N_B}$$

после перехода к пределу при $N \to \infty$ получаем *правило умножения вероятностей*:

$$\mathbf{P}\{A \text{ и } B\} = \mathbf{P}\{B\}\, \mathbf{P}\{A|B\} = \mathbf{P}\{A\}\, \mathbf{P}\{B|A\} \tag{6}$$

(последнее выражение получается после перемены ролей A и B). Итак, *вероятность совмещения двух событий равна произведению вероятности одного из них на условную вероятность другого в предположении, что первое событие произошло.*

Формула (6) особенно упрощается, если события A и B *независимые*; это значит, что информация о том, произошло или нет одно из этих событий, не влияет на вероятность другого, т. е.

$$\mathbf{P}\{A \mid B\} = \mathbf{P}\{A\}, \quad \mathbf{P}\{A \mid \overline{B}\} = \mathbf{P}\{A\}, \quad \mathbf{P}\{B \mid A\} = \mathbf{P}\{B\}, \quad \mathbf{P}\{B \mid \overline{A}\} = \mathbf{P}\{B\}.$$

(Впрочем, из равенств (1), (5) и (6) нетрудно вывести, что любое из этих четырех равенств влечет за собой остальные три.) Для независимых событий мы получаем

$$\mathbf{P}\{A \text{ и } B\} = \mathbf{P}\{A\}\,\mathbf{P}\{B\} \qquad (A \text{ и } B \text{ независимые}). \tag{7}$$

Эту формулу нетрудно распространить на случай *любого числа независимых событий*, т. е. таких, для которых информация о том, произошла или нет часть из этих событий, не влияет на вероятность оставшихся. Например,

$$\mathbf{P}\{A \text{ и } B \text{ и } C\} = \mathbf{P}\{A \text{ и } (B \text{ и } C)\} = \mathbf{P}\{A\}\,\mathbf{P}\{B \text{ и } C\} = \\ = \mathbf{P}\{A\}\,\mathbf{P}\{B\}\,\mathbf{P}\{C\}. \tag{8}$$

С помощью приведенных правил можно подсчитывать вероятности для ряда простых задач. Приведем пример. Пусть по мишени стреляют три стрелка, причем первый поражает ее с вероятностью 0,2, второй — с вероятностью 0,3, а третий — с вероятностью 0,5; какова вероятность того, что цель будет поражена хоть раз? Если обозначить через A_i ($i = 1, 2, 3$) событие, состоящее в поражении цели i-м стрелком, то нас интересует вероятность $\mathbf{P}\{A_1, \text{ или } A_2, \text{ или } A_3\}$. Пользоваться формулой (5) здесь нельзя, так как события A_i совместные: в цель могут попасть сразу два или три стрелка. Проще вычислить вероятность того, что все стрелки сделают промах, так как отдельные промахи независимы. В силу формул (1) и (8) получаем

$$\mathbf{P}\{A_1, \text{ или } A_2, \text{ или } A_3\} = 1 - \mathbf{P}\{\overline{A_1}, \text{ и } \overline{A_2}, \text{ и } \overline{A_3}\} = \\ = 1 - \mathbf{P}\{\overline{A_1}\}\,\mathbf{P}\{\overline{A_2}\}\,\mathbf{P}\{\overline{A_3}\} = 1 - 0{,}8 \cdot 0{,}7 \cdot 0{,}5 = 0{,}72.$$

Рассмотрим еще пример. Пусть из урны, в которой имеется три черных шара и один белый, дважды извлекается шар; какова вероятность того, что оба раза будет черный шар? Возможны два варианта этой задачи. Именно, если производится *выборка с возвращением*, т. е. первый шар после осмотра возвращается в урну и не исключено, что при второй выборке опять попадется он же, то по формуле (7) искомая вероятность равна $\frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$. Если же производится *выборка без возвращения*, т. е. первый шар после осмотра в урну не возвращается и потому во второй выборке не участвует, то искомая вероятность по формуле (6) равна $\frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}$.

5. Формула полной вероятности. Рассмотрим сначала пример. Пусть имеется три урны, причем в первой из них имеется три черных шара и один белый, во второй — один черный и три белых,

в третьей — только три черных шара. Мы наугад, с равной вероятностью, подходим к одной из урн, а затем наугад берем из нее шар; какова вероятность, что этот шар черный? Ясно, что если появился черный шар, то это значит, что либо мы подошли к первой урне и вынули из нее черный шар, либо подошли ко второй урне и сделали то же, либо подошли к третьей урне и сделали то же, причем эти три варианта попарно несовместимы. Вероятность первого варианта по формуле (7) равна $\frac{1}{3}\cdot\frac{3}{4}$, вероятность второго равна $\frac{1}{3}\cdot\frac{1}{4}$ и третьего $\frac{1}{3}\cdot 1$. Значит, вероятность того, что произойдет какой-либо из трех вариантов, по формуле (5) равна

$$\frac{1}{3}\cdot\frac{3}{4}+\frac{1}{3}\cdot\frac{1}{4}+\frac{1}{3}\cdot 1=\frac{2}{3}.$$

Это и есть искомая вероятность.

Рассмотрим теперь общий случай. Пусть в результате испытания должно произойти одно и только одно из некоторых событий B_1, B_2, ..., B_k (в примере это были подходы к одной из урн). Пусть, кроме того, рассматривается некоторое событие A (в примере это было вытягивание черного шара). Оно равносильно тому, что произойдет B_1 и A, либо B_2 и A, либо B_3 и A и т. д., причем все эти варианты попарно несовместны. Значит, по формуле (5)

$$\mathbf{P}\{A\}=\mathbf{P}\{(B_1 \text{ и } A), \text{ или } (B_2 \text{ и } A), \ \ldots, \text{ или } (B_k \text{ и } A)\}=$$
$$=\mathbf{P}\{B_1 \text{ и } A\}+\mathbf{P}\{B_2 \text{ и } A\}+\ldots+\mathbf{P}\{B_k \text{ и } A\}.$$

Отсюда по формуле (6) получаем окончательно

$$\mathbf{P}\{A\}=\mathbf{P}\{A|B_1\}\,\mathbf{P}\{B_1\}+\mathbf{P}\{A|B_2\}\,\mathbf{P}\{B_2\}+\ldots+\mathbf{P}\{A\,|\,B_k\}\,\mathbf{P}\{B_k\}. \quad (9)$$

Эта формула называется *формулой полной вероятности*. Она применяется к задачам, аналогичным разобранной в предыдущем абзаце.

6. Формула вероятностей гипотез. Начнем опять с тех же трех урн п. 5. Пусть мы наугад подошли к одной из урн, не зная их номеров, хотя и зная, что находится в первой урне, что — во второй, что — в третьей. Тогда если содержимое урн не проверяется, то все три гипотезы о том, что мы находимся у первой урны, у второй или у третьей, совершенно равноправны, т. е. вероятность каждой из них равна $\frac{1}{3}$. Но пусть мы наугад вытащили один шар из урны, у которой мы стоим, и он оказался белым. Тогда указанные гипотезы придется пересмотреть: например, ясно, что наша урна не может быть третьей, кроме того, вероятней, что она вторая, а не первая (почему?). Первоначальные вероятности гипотез называются доопытными, *априорными* (от латинского a priori, изначально), а пересмотренные — *апостериорными* (a posteriori, из последующего). Как же найти эти пересмотренные вероятности?

Рассмотрим общий случай. Пусть имеется несколько гипотез H_1, H_2, ..., H_k, причем известно, что на самом деле выполняется одна и только одна из них и априорные вероятности этих гипотез равны соответственно $\mathbf{P}\{H_1\}$,

$\mathbf{P}\{H_2\}, \ldots, \mathbf{P}\{H_k\}$. Пусть известны вероятности $\mathbf{P}\{A|H_i\}$ $(i=1, 2, \ldots, k)$ некоторого события A при каждой из этих гипотез. Тогда апостериорные вероятности — это как раз вероятности $\mathbf{P}\{H_i|A\}$. Чтобы их найти, напишем на основе (6)

$$\mathbf{P}\{A\}\mathbf{P}\{H_i|A\}=\mathbf{P}\{H_i\}\mathbf{P}\{A|H_i\},$$

откуда с помощью формулы (9) получим

$$\mathbf{P}\{H_i|A\}=\frac{\mathbf{P}\{A|H_i\}\mathbf{P}\{H_i\}}{\mathbf{P}\{A|H_1\}\mathbf{P}\{H_1\}+\mathbf{P}\{A|H_2\}\mathbf{P}\{H_2\}+\ldots+\mathbf{P}\{A|H_k\}\mathbf{P}\{H_k\}}.$$

Это и есть искомая *формула вероятностей гипотез* или *формула Бáйеса*. Мы предоставляем читателю убедиться с ее помощью, что в задаче предыдущего абзаца пересмотренные вероятности равны соответственно $\frac{1}{4}$, $\frac{3}{4}$ и 0.

7. Принцип игнорирования маловероятных событий. Мы видим, что методы теории вероятностей дают возможность вычислять вероятности различных событий. Как же пользоваться этими результатами? Вряд ли кого-нибудь удовлетворит такой ответ: данное событие либо произойдет, либо нет.

Выход из этого затруднения типичен для прикладной математики. Он состоит в том, что если вероятность рассматриваемого события A окажется достаточно малой, $\mathbf{P}\{A\}<\varepsilon$, где ε — достаточно малое число, то просто полагают $\mathbf{P}\{A\}=0$, т. е. считают событие A *практически невозможным*, возможность A игнорируют. Конечно, в принципе A не исключено, поэтому остается возможность того, что предсказание о ненаступлении A окажется ошибочным; однако чем ε меньше, тем эта возможность будет появляться все реже.

Какое же все-таки взять ε? В различных областях приложений математики здесь существуют свои традиции. Если нет никакой опасности в том, что событие A произойдет, т. е. ошибка от неправильного предсказания легко исправима, то можно принять $\varepsilon=0{,}1$. Это означает, что в среднем из десяти прогнозов один будет неверным. Однако если достижение большей надежности не связано с существенными осложнениями, то чаще полагают $\varepsilon=0{,}01$. Это означает примерно, что игнорируется возможность падения монеты гербом кверху семь раз подряд, так как $2^7\approx 100$. Для еще более точных прогнозов можно принять $\varepsilon=0{,}001$, тогда в среднем будет один ошибочный прогноз из тысячи и т. д. Чем меньше взять ε, тем предсказание будет точнее, но тем сложнее обеспечить эту точность. Особенно высокий критерий точности надо взять, если ошибочный прогноз может быть связан с человеческими жертвами; при этом во многих случаях приходится переходить от вероятностных оценок к детерминированным.

В п. 14 будут указаны более основательные соображения для выбора критерия маловероятности, т. е. ε.

§ 2. Случайные величины

8. Определения. *Случайной величиной называется величина, которая в результате испытания принимает определенное значение*; при повторных испытаниях эти значения, вообще говоря, различные.

Случайными являются количество студентов на лекции, длина изделия из какой-либо партии изделий, продолжительность человеческой жизни и т. п.

Как и любая величина (п. I.5), случайная величина может быть *непрерывной* или *дискретной*. Так, первая из приведенных выше случайных величин дискретна, остальные две непрерывны. При этом существенно, что возможные значения количества студентов являются дискретными (целыми) до испытаний, тогда как дискретные значения для длин изделий до испытаний задать нельзя.

Чтобы задать дискретную случайную величину, надо перечислить ее возможные значения и вероятности, с которыми они достигаются. Получается таблица такого вида:

Дискретная случайная величина ξ

значения	x_1	x_2	x_3	. . . .
вероятности	P_1	P_2	P_3	. . . .

(10)

Эта таблица может быть конечной или бесконечной. При этом все вероятности P_k должны быть положительными, а их сумма в силу формулы (2) должна равняться единице. В крайнем случае, если имеется всего одно возможное значение, оно должно достигаться с вероятностью 1, т. е. обязательно; это — случай *детерминированной* величины.

Непрерывная случайная величина ξ может принимать все значения. Однако вероятность того, что такая величина будет в точности равняться любому заданному значению x, равна нулю. (Подобно тому как при рассмотрении непрерывно распределенной массы масса любой отдельной точки считается равной нулю.) Можно говорить о вероятности попадания величины ξ на заданный интервал оси x. При этом вероятность того, что ξ примет значения из бесконечно малого интервала от x до $x+dx$, бесконечно мала, пропорциональна dx и зависит от x; таким образом, эта вероятность равна $p(x)\,dx$, где функция $p(x)$ — *плотность вероятности* (*плотность распределения*) — полностью характеризует случайную величину ξ. Ясно, что $p(x) \geqslant 0$ $(-\infty < x < \infty)$. Из формулы (5) вытекает, что вероятность попадания случайной величины ξ на интер-

вал a, b равна $\int_a^b p(x)\,dx$, а из формулы (2) следует, что

$$\int_{-\infty}^{\infty} p(x)\,dx = 1. \tag{11}$$

Впредь интеграл от $-\infty$ до ∞ мы будем записывать без обозначения пределов интегрирования, так что формула (11) *будет иметь вид* $\int p(x)\,dx = 1$. Неопределенных интегралов в этой главе не будет. (Укажем по этому поводу, что вообще в приложениях математики знак $\int$ для неопределенного интеграла применяется весьма редко. Обычно $\int$ означает определенный интеграл, для которого область интегрирования ясна из контекста, например интегрирование производится по максимальному диапазону изменения переменных интегрирования.)

Если пользоваться дельта-функцией (п. XIV.25), то можно говорить о плотности вероятности и для дискретной случайной величины. Если эта величина задается таблицей (10), то

$$p(x) = P_1\delta(x-x_1) + P_2\delta(x-x_2) + P_3\delta(x-x_3) + \ldots$$

С помощью дельта-функции можно рассматривать и случайные величины смешанного непрерывно-дискретного типа. В общем случае задание случайной величины равносильно заданию неотрицательной меры на прямой (п. XVI.19), причем такой, что мера всей прямой равна единице. При этом мера любого интервала есть вероятность того, что случайная величина попадет в этот интервал.

9. Примеры дискретных случайных величин. Пусть рассматривается случайное событие A, причем $\mathbf{P}\{A\} = P$. Пусть будет проведено одно испытание, сколько раз произойдет событие A? Ясно, что либо 1, либо 0, т. е. получается случайная величина, принимающая два значения: 1 с вероятностью P и 0 с вероятностью $1-P$ (см. формулу (1)).

Пусть теперь испытание будет проведено несколько раз, например, три раза и событие A произойдет ν раз. Тогда ν — случайная величина, которая может принимать значения 0, 1, 2 и 3. Подсчитаем вероятности этих значений. Если $\nu = 0$, то все три раза событие A не произошло. Вероятность этого в силу независимости испытаний, согласно формуле (8), равна $(1-P)^3$. Значение $\nu = 1$ получается в трех вариантах: если событие A произойдет при первом (втором, третьем) испытании и не произойдет в двух других. Вероятность каждого из этих вариантов в силу формулы (8) равна $P(1-P)^2$, а потому вероятность того, что встретится какой-либо один из них, по формуле (5) равна $3P(1-P)^2$. Аналогично разбираются значения $\nu = 2$ и $\nu = 3$ и мы приходим к таблице:

значения ν	0	1	2	3
вероятности	$(1-P)^3$	$3P(1-P)^2$	$3P^2(1-P)$	P^3

(Проверьте, что сумма полученных вероятностей равна единице!)

Аналогично разбирается общий случай n испытаний. Событие A произойдет ν раз, где ν — случайная величина, распределенная по *биномиальному закону:*

значения ν	0	1	2	...	n
вероятности	$(1-P)^n$	$\binom{n}{1} P(1-P)^{n-1}$	$\binom{n}{2} P^2(1-P)^{n-2}$	...	P^n

Здесь $\binom{n}{k}=\frac{n(n-1)\ldots(n-k+1)}{k!}$ — биномиальный коэффициент, равный числу вариантов, при которых событие A произойдет ровно k раз, т. е. числу сочетаний из n элементов по k.

Пример Монета бросается шесть раз, какова вероятность того, что она упадет кверху гербом ровно три раза?

Ответ: $\binom{6}{3}\frac{1}{2^3}\left(1-\frac{1}{2}\right)^3=\frac{5}{16}$.

Рассмотрим теперь, что получится из биномиального закона, если число n испытаний весьма велико, а вероятность P события A весьма мала, причем так, что произведение $\alpha=Pn$ остается конечным. Совершая в формуле

$$\mathbf{P}\{\nu=k\}=\binom{n}{k}P^k(1-P)^{n-k}=\frac{n(n-1)\ldots(n-k+1)}{k!}\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^n$$

переход к пределу при $n\to\infty$, что мы предоставим читателю, получим в пределе

$$\mathbf{P}\{\nu=k\}=\frac{\alpha^k}{k!}e^{-\alpha} \quad (k=0,\ 1,\ 2,\ \ldots).$$

Таким образом, получается случайная величина, принимающая бесконечное количество значений:

значения ν	0	1	2	...	k	...
вероятности	$e^{-\alpha}$	$\frac{\alpha}{1!}e^{-\alpha}$	$\frac{\alpha^2}{2!}e^{-\alpha}$	...	$\frac{\alpha^k}{k!}e^{-\alpha}$	...

Этот закон распределения называется *законом Пуассона* по имени выдающегося французского механика, физика и математика С. Пуассóна (1781 — 1840).

Типичным примером величины, распределенной по закону Пуассона, может служить число атомов некоторой порции долгоживущего радиоактивного вещества, распадающихся за промежуток времени, выбранный так, чтобы можно было регистрировать распад отдельных атомов. Этот закон получается из-за того, что отдельные атомы распадаются с равной вероятностью независимо друг от друга. Имеется и ряд других подобных примеров.

10. Примеры непрерывных случайных величин. Одним из самых простых примеров служит величина, *равномерно распределенная* по некоторому интервалу a, b, т. е. принимающая все значения из этого интервала с равной вероятностью и не принимающая значений вне этого интервала. Плотность вероятности такой величины равна

$$p(x)=\begin{cases} c & (a\leqslant x\leqslant b),\\ 0 & (x<a \text{ или } x>b).\end{cases}$$

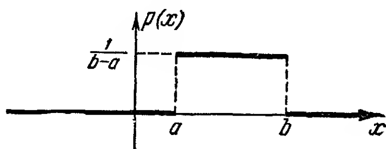


Рис 343.

Из условия (11) получаем, что $c=\frac{1}{b-a}$. График этой плотности показан на рис. 343. Этому закону подчинена, например, ошибка, получающаяся от округления числовой величины до ее ближайшего целого значения, в этом случае $a=-0{,}5$, $b=0{,}5$, $c=1$ (почему?).

Наиболее распространены величины, распределенные по *нормальному закону* или, что то же, *закону Гаусса*. Плотность распределения такой величины выражается формулой

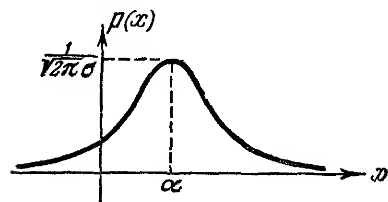


Рис. 344.

$$p(x)=Me^{-\beta(x-\alpha)^2}=M\exp\left[-\beta(x-\alpha)^2\right],$$

где α, $M>0$ и $\beta>0$ — числовые параметры. Из них M легко выразить через β с помощью формулы (11), подстановки $\sqrt{\beta}(x-\alpha)=s$ и применения интеграла (XIV.72), что даст (проверьте!) $M=\sqrt{\frac{\beta}{\pi}}$. Кроме того, по причинам, которые будут ясны из п. 15, принято обозначать $\beta=\frac{1}{2\sigma^2}$, т. е.

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\alpha)^2}{2\sigma^2}\right]. \qquad (12)$$

График этой плотности показан на рис. 344. О причинах распространенности закона Гаусса будет сказано в § 4.

Вероятность попадания величины ξ, распределенной по закону (12), на интервал a, b равна

$$\mathbf{P}\{a\leqslant\xi\leqslant b\}=\frac{1}{\sqrt{2\pi}\,\sigma}\int_a^b \exp\left[-\frac{(x-\alpha)^2}{2\sigma^2}\right]dx. \tag{13}$$

Она легко выражается через *интеграл вероятностей*

$$\Phi(t)=\sqrt{\frac{2}{\pi}}\int_0^t \exp\left(-\frac{s^2}{2}\right)ds,$$

для которого имеются подробные таблицы (см., например, [9], [40], [53]). В самом деле, подставляя $\frac{(x-\alpha)}{\sigma}=s$ в (13), получаем

$$\mathbf{P}\{a\leqslant\xi\leqslant b\}=\frac{1}{\sqrt{2\pi}}\int_{(a-\alpha)/\sigma}^{(b-\alpha)/\sigma}\exp\left(-\frac{s^2}{2}\right)ds=\frac{1}{2}\sqrt{\frac{2}{\pi}}\left(\int_0^{(b-\alpha)/\sigma}-\int_0^{(a-\alpha)/\sigma}\right)=$$

$$=\frac{1}{2}\left[\Phi\left(\frac{b-\alpha}{\sigma}\right)-\Phi\left(\frac{a-\alpha}{\sigma}\right)\right]. \tag{14}$$

11. Системы случайных величин. Мы ограничимся более важным случаем непрерывных величин, причем для простоты будем рассматривать систему из двух величин. Дискретные величины и системы из большего числа величин исследуются аналогично.

Пусть рассматриваются одновременно две случайные величины ξ, η, которые в результате одного и того же испытания принимают определенные значения. Тогда вероятность того, что ξ окажется между x и $x+dx$, а η — между y и $y+dy$, должна быть пропорциональна как dx, так и dy, т. е. эта вероятность равна $p(x, y)\,dx\,dy$. Функция $p(x, y)$ называется *плотностью совместного распределения* величин ξ, η. Эта функция полностью определяет пару случайных величин ξ, η. Ясно, что она удовлетворяет условиям

$$p(x, y)\geqslant 0, \quad \int dx\int p(x, y)\,dy=1.$$

Зная плотность совместного распределения, легко подсчитать плотности распределения каждой из величин ξ, η в отдельности. В самом деле, в силу формулы (5) вероятность того, что ξ примет значения между x и $x+dx$, а η будет какой угодно, равна

$$\mathbf{P}\{x\leqslant\xi\leqslant x+dx\}=\int_{y=-\infty}^{\infty}p(x, y)\,dx\,dy=\left(\int p(x, y)\,dy\right)dx.$$

В силу п. 9 отсюда получаем, что плотность распределения величины ξ, $p_\xi(x)=\int p(x, y)\,dy$. Аналогично выводим, что $p_\eta(y)=$ $=\int p(x, y)\,dx$. Обратно, зная только $p_\xi(x)$ и $p_\eta(y)$, в общем слу-

чае восстановить функцию $p(x, y)$ нельзя, так как существенно «взаимодействие» величин ξ, η.

Имеется важный частный случай, когда такое восстановление возможно. Это — случай, когда величины ξ, η являются *независимыми*, т. е. когда информация о значениях одной из них не влияет на вероятности значений другой. Тогда в силу формулы (7)

$$p(x, y)\,dx\,dy = \mathbf{P}\{x \leqslant \xi \leqslant x+dx, \quad y \leqslant \eta \leqslant y+dy\} = \\ = \mathbf{P}\{x \leqslant \xi \leqslant x+dx\}\,\mathbf{P}\{y \leqslant \eta \leqslant y+dy\} = p_\xi(x)\,dx \cdot p_\eta(y)\,dy,$$

т. е.

$$p(x, y) = p_\xi(x)\,p_\eta(y). \tag{15}$$

Обратно, если это условие выполнено, то легко проверить, что величины ξ, η независимые.

С системами случайных величин связано понятие *многомерной случайной* величины, принимающей значения из некоторого многомерного пространства (п. X.2) (R). Закон распределения такой величины задается некоторой неотрицательной мерой в (R) (п. XVI.19), причем мера всего (R) должна равняться единице, при этом мера какой-либо области просто равна вероятности того, что рассматриваемая величина попадет в эту область. Если в (R) имеется некоторая «каноническая» мера, как мера Лебега в конечномерном евклидовом пространстве, то с помощью дифференцирования по этой мере (п XVI.7) можно перейти к плотности вероятности. Если в (R) ввести обобщенные координаты $t_1, t_2, \ldots, t_k$, то можно взамен многомерной случайной величины рассматривать совокупность этих координат как систему обычных случайных величин.

В качестве простого примера приведем *многомерный закон Гаусса*. Это — закон распределения случайного вектора $\boldsymbol{\xi}$ в пространстве E_n (п. VII.18) с плотностью распределения

$$p(\mathbf{x}) = M \exp(-\mathbf{x}^*\mathbf{A}\mathbf{x}), \tag{16}$$

где $\mathbf{A}$ — некоторая симметрическая положительно определенная матрица (пп. XI.11 и XII.7), а M — *нормировочный коэффициент*, необходимый для равенства интеграла от $p(\mathbf{x})$ единице. Как известно из п. XI.11, после введения нового декартова базиса в E_n функция (16) приобретает вид

$$p(\mathbf{x}') = M \exp(-\lambda_1 x_1'^2 - \lambda_2 x_2'^2 - \ldots - \lambda_n x_n'^2) = \\ = M \exp(-\lambda_1 x_1'^2) \exp(-\lambda_2 x_2'^2) \ldots \exp(-\lambda_n x_n'^2),$$

где $\lambda_1, \lambda_2, \ldots, \lambda_n$ — собственные значения матрицы $\mathbf{A}$. Отсюда легко вычислить, что $M = \sqrt{\lambda_1\lambda_2\ldots\lambda_n}\,\pi^{-n/2}$, кроме того, из формулы (15) видно, что случайные координаты в новом базисе являются независимыми.

12. Функции от случайных аргументов. Если $\eta = f(\xi)$, где ξ — случайная величина, то и η будет случайной величиной, причем если ξ дискретна (непрерывна), то и η дискретна (соответственно непрерывна). Если ξ определяется таблицей (10), то η будет принимать значение $f(x_1)$ с вероятностью p_1, значение $f(x_2)$ с вероятностью p_2 и т. д.; при этом надо иметь в виду, что если $f(x_i) = f(x_j)$, то соответствующие вероятности p_i и p_j, конечно, складываются. Например, если ξ принимает значения $-2, -1, 0, 1, 2$

с равной вероятностью $\frac{1}{5}$, то ξ^2 принимает значения 0, 1 и 4 с вероятностями $\frac{1}{5}$, $\frac{2}{5}$ и $\frac{2}{5}$ (почему?).

Если $\eta=f(\xi)$ и ξ — непрерывная случайная величина с плотностью распределения $p_\xi(x)$, то если функция $f(\xi)$ возрастающая, плотность распределения $p_\eta(y)$ равна

$$p_\eta(y)=\frac{1}{dy}\mathbf{P}\{y\leqslant\eta\leqslant y+dy\}=\frac{1}{dy}\mathbf{P}\{x\leqslant\xi\leqslant x+dx\}=\frac{1}{dy}p_\xi(x)\,dx=\frac{p_\xi(x)}{f'(x)},$$

где в правой части x находится из равенства $f(x)=y$. Если функция $f(\xi)$ убывающая, то в правой части взамен $f'(x)$ надо поставить $|f'(x)|$. Наконец, если функция $f(x)$ не монотонная, то взамен правой части надо поставить сумму аналогичных выражений, взятую по всем решениям уравнения $f(x)=y$.

Аналогично рассматриваются функции от нескольких случайных переменных. Пусть, например, $\zeta=f(\xi,\eta)$, где пара случайных величин характеризуется совместной функцией распределения $p(x, y)$. Тогда

$$p_\zeta(z)=\frac{1}{dz}\mathbf{P}\{z\leqslant\zeta\leqslant z+dz\}=\frac{d}{dz}\iint\limits_{f(x,\,y)\leqslant z}p(x,\,y)\,dx\,dy. \tag{17}$$

Эта производная в общем случае вычисляется по методам п. XVI.18.

Покажем применение этого результата к вычислению плотности распределения $p_\zeta(z)$ суммы $\zeta=\xi+\eta$ двух независимых случайных величин. В силу формул (15) и (17) получаем (проверьте!)

$$p_\zeta(z)=\frac{d}{dz}\iint\limits_{x+y\leqslant z}p_\xi(x)\,p_\eta(y)\,dx\,dy=\frac{d}{dz}\int dx\int\limits_{-\infty}^{z-x}p_\xi(x)\,p_\eta(y)\,dy=$$

$$=\frac{d}{dz}\int\left[\int\limits_{-\infty}^{z-x}p_\eta(y)\,dy\right]p_\xi(x)\,dx=\int\left[\frac{d}{dz}\int\limits_{-\infty}^{z-x}p_\eta(y)\,dy\right]p_\xi(x)\,dx=$$

$$=\int p_\xi(x)\,p_\eta(z-x)\,dx.$$

Мы предлагаем читателю в качестве упражнения доказать, что сумма двух независимых слагаемых, каждое из которых равномерно распределено на интервале $0\leqslant x\leqslant 1$, имеет плотность распределения

$$p(x)=\begin{cases}0 & (x\leqslant 0,\ x\geqslant 2),\\ x & (0\leqslant x\leqslant 1),\\ 2-x & (1\leqslant x\leqslant 2).\end{cases}$$

§ 3. Числовые характеристики случайных величин

13. Среднее значение. Пусть имеется дискретная случайная величина ξ, определенная таблицей (10), и пусть произведено большое число N испытаний; каково будет среднее арифметическое полученных значений ξ? Обозначим через N_i число случаев, когда ξ при

этих испытаниях равнялось x_i. Тогда искомое среднее арифметическое равно

$$\frac{N_1x_1+N_2x_2+N_3x_3+\ldots}{N}=x_1\frac{N_1}{N}+x_2\frac{N_2}{N}+x_3\frac{N_3}{N}+\ldots$$

Однако мы знаем (п. 2), что в пределе, при $N\longrightarrow\infty$, будет $\frac{N_i}{N}\longrightarrow P_i$. Значит, в пределе получается значение

$$x_1P_1+x_2P_2+x_3P_3+\ldots \tag{18}$$

Оно называется *средним значением* (*центром распределения, математическим ожиданием*) случайной величины ξ и является ее важнейшей характеристикой; оно обозначается буквами $\bar{\xi}$, $\mathbf{M}\{\xi\}$, $\mathbf{M}\xi$. Среднее значение случайной величины есть величина неслучайная, детерминированная. (Проверьте, например, что среднее значение очков, выпадающих при бросании игрального кубика, равно 3,5.)

Если ξ — непрерывная случайная величина с плотностью распределения $p(x)$, то взамен (18) получится формула

$$\bar{\xi}=\sum x\,dP=\sum xp(x)\,dx=\int xp(x)\,dx. \tag{19}$$

(Знак суммы написан только для аналогии, на самом деле это, конечно, интеграл.) Мы предоставляем читателю проверить по этой формуле, что средние значения величин, рассмотренных в п. 10, равны соответственно $\frac{a+b}{2}$ и α. Это, впрочем, ясно и из соображений симметрии.

14. Свойства среднего значения. 1. Из определения очевидно, что *среднее значение случайной величины ξ имеет ту же размерность, что и ξ, и заключено между наименьшим и наибольшим возможными значениями ξ.*

2. *Если случайную величину умножить на константу*, т. е. детерминированную величину, *то среднее значение умножится на ту же константу*: $\mathbf{M}\{C\xi\}=C\mathbf{M}\{\xi\}$ $(C=\text{const})$. Это вытекает из п. 13, так как при умножении всех значений на константу и среднее арифметическое умножится на ту же константу. Аналогично доказывается следующее свойство:

3. *Если сложить две случайные величины, то их средние значения также сложатся*: $\mathbf{M}\{\xi+\eta\}=\mathbf{M}\{\xi\}+\mathbf{M}\{\eta\}$. В частности, *если к случайной величине добавить константу, то к среднему значению добавится та же константа.*

С помощью этого свойства легко, в частности, получить значение среднего значения величины ξ, распределенной по биномиальному закону (п. 9). В самом деле,

$$\xi=\xi_1+\xi_2+\ldots+\xi_n, \tag{20}$$

где $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, принимающие значение 1 с вероятностью P и значение 0 с вероятностью $1-P$ (ξ_i указывает, сколько раз событие A произошло при i-м испытании). Отсюда $\mathbf{M}\{\xi\}=\mathbf{M}\{\xi_1\}+\mathbf{M}\{\xi_2\}+\ldots+\mathbf{M}\{\xi_n\}=nP$. Впрочем, это легко понять и непосредственно. Из полученного результата вытекает, что для распределения Пуассона (п. 9) $\mathbf{M}\{\xi\}=\alpha$.

4. *Среднее значение произведения двух независимых случайных величин равно произведению их средних значений:* $\mathbf{M}\{\xi\eta\}=\mathbf{M}\{\xi\}\,\mathbf{M}\{\eta\}$, если ξ и η независимы. В самом деле, если ξ принимают значения x_i с вероятностями p_i, а η — значения y_j с вероятностями q_j, то $\xi\eta$ принимает значения x_iy_j, вероятности которых в силу независимости ξ и η равны p_iq_j (п. 4). Поэтому

$$\mathbf{M}\{\xi\eta\}=\sum_{i,j}x_iy_j(p_iq_j)=\sum_i\Big(\sum_j x_iy_jp_iq_j\Big)=\\ =\sum_i x_ip_i\sum_j y_jq_j=\mathbf{M}\{\xi\}\,\mathbf{M}\{\eta\}.$$

Свойства 3 и 4 немедленно распространяются на любое число слагаемых и сомножителей. Подчеркнем, что в свойстве 4 существенна независимость сомножителей; если от нее отказаться, то утверждение, вообще говоря, будет неверным. Например, если случайную величину умножить саму на себя, то среднее значение, как правило, вовсе не возведется в квадрат, т. е. $\overline{\xi^2}\neq\bar{\xi}^2$: так, в примере начала п. 12 было $\mathbf{M}\{\xi\}=0$, $\mathbf{M}\{\xi^2\}=2$ (проверьте!).

5. *Если случайная величина ξ принимает симметричные относительно некоторой константы a значения с равной вероятностью, то* $\bar{\xi}=a$ (очевидно).

6. Если случайная величина ξ характеризуется таблицей (10), то $\overline{f(\xi)}=\sum_i f(x_i)\,p_i$; *если случайная величина ξ имеет плотность распределения $p(x)$, то* $\overline{f(\xi)}=\int f(x)\,p(x)\,dx$ (сразу следует из определений).

Подсчет среднего значения дает, в частности, возможность подойти к выбору критерия маловероятности (п. 7). Мы приведем здесь самые простые соображения по этому поводу. Именно, допустим, что события с вероятностью, меньшей некоторого ε, считаются невозможными, игнорируются. Тогда возможен неправильный прогноз, т. е. событие, объявленное невозможным, на самом деле произойдет. Пусть ущерб от неправильного прогноза равен k рублей, так что средний ущерб равен εk. Поэтому желательно уменьшить ε, но повышение качества прогнозов требует расходов. Обозначим через $f(\varepsilon)$ расход в рублях на одно испытание, необходимое для обеспечения прогноза «с точностью до ε»; эта функция, примерный график которой показан на рис. 345, в конкретных задачах может быть ориентировочно составлена. Таким образом, средний расход, связанный с неправильными прогнозами, равен $f(\varepsilon)+k\varepsilon$; значение $\varepsilon=\varepsilon_0$ надо выбирать из соображения минимизации этой суммы.

Скажем в заключение о многомерных случайных величинах (п. 11), принимающих значения в конечномерном линейном пространстве (R) (п. VII. 17), т. е. о случайных векторных величинах. Формула для среднего значения такой величины выглядит аналогично (19):

$$\bar{\xi}=\int\limits_{(R)} \mathrm{x}\,dP=\int\limits_{(R)} \mathrm{x}p(\mathrm{x})\,dR,$$

где интегрирование выполняется по всему пространству (R). Все свойства среднего значения аналогичны скалярному случаю. При этом свойство 4 имеет место для всех видов произведений, для которых имеют место обычные правила раскрытия скобок (произведение вектора на скаляр, скалярное или векторное произведения векторов).

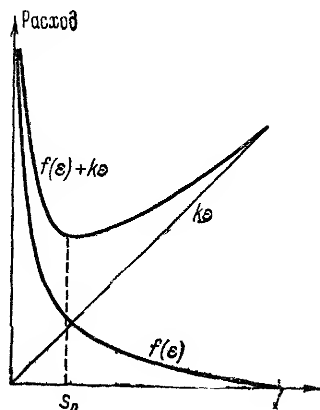


Рис. 345.

15. Дисперсия. Дисперсия характеризует степень рассеянности случайной величины вокруг своего среднего значения. Пусть дана случайная величина ξ; тогда по определению ее *дисперсией* называется величина

$$\mathbf{D}\xi=\mathbf{D}\{\xi\}=\mathbf{M}(\xi-\mathbf{M}\xi)^2. \qquad (21)$$

Это — величина детерминированная и всегда положительная, за исключением случая, когда сама ξ детерминированная, тогда $\mathbf{D}\xi=0$.

Из (21) в силу свойства 6 п. 14 вытекают формулы

$$\mathbf{D}\xi=\sum_i (x_i-\bar{\xi})^2 P_i, \quad D\xi=\int (x-\bar{\xi})^2 p(x)\,dx.$$

Из определения (21) легко следует, что *если ξ умножить на константу C, то $\mathbf{D}\xi$ умножится на C^2; если к ξ прибавить константу, то $\mathbf{D}\xi$ не изменится*. Далее, *если случайные величины ξ и η независимые, то*

$$\mathbf{D}\{\xi+\eta\}=\mathbf{D}\xi+\mathbf{D}\eta. \qquad (22)$$

В самом деле,

$$\mathbf{D}\{\xi+\eta\}=\mathbf{M}\{[\xi+\eta-\mathbf{M}(\xi+\eta)]^2\}=\mathbf{M}\{[(\xi-\mathbf{M}\xi)+(\eta-\mathbf{M}\eta)]^2\}=$$
$$=\mathbf{M}(\xi-\mathbf{M}\xi)^2+2\mathbf{M}\{(\xi-\mathbf{M}\xi)(\eta-\mathbf{M}\eta)\}+\mathbf{M}(\eta-\mathbf{M}\eta)^2=$$
$$=\mathbf{D}\xi+2\mathbf{M}(\xi-\mathbf{M}\xi)\cdot\mathbf{M}(\eta-\mathbf{M}\eta)+\mathbf{D}\eta=\mathbf{D}\xi+2\cdot0\cdot0+\mathbf{D}\eta=\mathbf{D}\xi+\mathbf{D}\eta$$

(где использована независимость величин ξ, η?).

Подсчитаем дисперсии в примерах пп. 9—10. Дисперсия каждого слагаемого в формуле (20) равна $(0-P)^2(1-P)+(1-P)^2P=$ $=P(1-P)$. Значит, в силу формулы (22) для биномиального закона $\mathbf{D}\xi=nP(1-P)$. Отсюда, переходя к пределу при $n\to\infty$, получаем,

что для закона Пуассона $\mathbf{D}\xi=\alpha$. Для равномерного распределения

$$\mathbf{D}\xi=\int\left(x-\frac{a+b}{2}\right)^2 p(x)\,dx=\int_a^b\left(x-\frac{a+b}{2}\right)^2\frac{1}{b-a}\,dx=\frac{(b-a)^2}{12}$$

(проверьте!). Наконец, для закона Гаусса

$$\mathbf{D}\xi=\int(x-\alpha)^2\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\alpha)^2}{2\sigma^2}\right]dx=\left|\frac{x-\alpha}{\sqrt{2}\sigma}=s\right|=$$

$$=\frac{2\sigma^2}{\sqrt{\pi}}\int s^2\exp(-s^2)\,ds=\left|\begin{matrix} s=u & s\exp(-s^2)=dv \\ ds=du & -\frac{1}{2}\exp(-s^2)=v\end{matrix}\right|=$$

$$=\frac{\sigma^2}{\sqrt{\pi}}\int\exp(-s^2)\,ds=\sigma^2.$$

Часто наряду с дисперсией $\mathbf{D}\xi$ пользуются квадратным корнем из нее, называемым *средним квадратичным отклонением* величины ξ, так как $\sqrt{\mathbf{D}\xi}$ имеет ту же размерность, что и ξ. Мы видим, что σ в законе Гаусса (12) — это среднее квадратичное отклонение.

Из формулы (22) вытекает важное следствие. Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и распределены по одинаковому закону со средним квадратичным отклонением σ. Тогда их сумма имеет дисперсию $n\sigma^2$ и потому среднее квадратичное отклонение $\sqrt{n}\sigma$. Однако, как видно из двух последних примеров и будет подробно разобрано в п. 19, основные значения случайной величины группируются на интервале, длина которого пропорциональна среднему квадратичному отклонению. Значит, у суммы n независимых слагаемых длина этого интервала пропорциональна не n (как было бы для равных слагаемых), а $\sqrt{n}$. Этому закону подчиняется, в частности, погрешность суммы слагаемых, известных с одинаковой точностью (ср. п. I.9).

Две случайные величины, распределенные по различным законам, могут иметь соответственно равные средние значения и дисперсии. Поэтому в более полных исследованиях применяются и другие числовые, детерминированные характеристики случайных величин. В частности, применяются *моменты распределения*

$$\mathbf{M}\{\xi^k\}=\int x^k p(x)\,dx \qquad (k=1,\ 2,\ 3,\ \ldots);$$

конечно, предполагается, что интеграл сходится. Первый момент — это как раз среднее значение, а дисперсия в силу формулы (21) выражается через второй момент:

$$\mathbf{D}\xi=\mathbf{M}\{\xi^2\}-2\mathbf{M}\{\xi\mathbf{M}\xi\}+\mathbf{M}\{(\mathbf{M}\xi)^2\}=\mathbf{M}\{\xi^2\}-(\mathbf{M}\xi)^2.$$

Моменты высших порядков более полно характеризуют закон распределения рассматриваемой случайной величины.

16. Корреляционная зависимость. Пусть задана совместная плотность распределения $p(x, y)$ двух случайных величин ξ, η (п. 11). Тогда задание любого значения $\xi=a$ определяет *условное распределение вероятностей* с плотностью $p(a, y)$ величины η. При изменении значения ξ закон условного распределения величины η, вообще говоря, меняется. Такая зависимость величины η от величины ξ называется *корреляционной*, в отличие от детерминированной, т. е. функциональной зависимости, которая изучалась в предыдущих

главах. Аналогично задание совместного распределения какого угодно числа случайных величин определяет корреляционную зависимость любой из этих величин от остальных

Корреляционные зависимости встречаются очень часто. Например, когда мы говорим о зависимости веса человека от его роста, то речь идет, конечно, о корреляционной зависимости, так как вес не определяется ростом однозначно, в то же время ясно, что закон распределения весов двухметровых людей совсем не тот, что полутораметровых. Когда мы говорим, что курение сокращает продолжительность жизни, то, конечно, речь также идет о корреляционной зависимости, так как имеются и противоположные примеры; однако закон распределения продолжительности жизни у некурящих такой, что среднее значение этой продолжительности выше, чем у курящих. Распознание корреляционных зависимостей требует гораздо большего внимания, чем функциональных, так как на многих людей противоречащие примеры производят слишком большое впечатление. (На непонимании разницы между функциональной и корреляционной зависимостями основано известное порицание: «А еще в очках!»)

Среднее значение условного распределения величины η будет уже детерминированной функцией от ξ:

$$M\{\eta \mid \xi = x\} = \int y p(x, y)\, dy,$$

которую мы обозначим через $f(x)$. Она называется *функцией регрессии величины η на величину ξ*; в примере с продолжительностью жизни, конечно, надо говорить именно о функции регрессии. Аналогично определяется функция регрессии $\varphi(y)$ величины ξ на величину η; интересно, что функции $f(x)$ и $\varphi(y)$, вообще говоря, не являются взаимно обратными, как это будет для детерминированной зависимости. Это особенно ясно для независимых случайных величин ξ, η, когда обе функции — константы.

Сравнительно прост случай *линейной корреляции*, для которого обе функции, $f(x)$ и $\varphi(y)$, линейные. Обозначим

$$r_{\xi,\eta} = \frac{\overline{\xi\eta} - \bar{\xi}\bar{\eta}}{\sqrt{\mathrm{D}\xi\, \mathrm{D}\eta}}$$

— это *коэффициент корреляции* между величинами ξ и η. Можно доказать, что всегда $|r_{\xi,\eta}| \leqslant 1$ и что функции регрессии η на ξ и ξ на η в линейном случае имеют вид

$$f(x) = r_{\xi,\eta} \sqrt{\frac{\mathrm{D}\eta}{\mathrm{D}\xi}}(x - \bar{\xi}) + \bar{\eta}, \qquad \varphi(y) = r_{\xi,\eta} \sqrt{\frac{\mathrm{D}\xi}{\mathrm{D}\eta}}(y - \bar{\eta}) + \bar{\xi};$$

мы не будем на этом останавливаться. Отсюда видно, что при $r_{\xi,\eta} > 0$ обе функции являются возрастающими, а при $r_{\xi,\eta} < 0$ — убывающими.

Важный частный случай линейной корреляции дает двумерный закон Гаусса (п 11) с плотностью распределения $p(x, y) = M \exp[-(Ax^2 + 2Bxy + Cy^2)]$, где M — нормировочный коэффициент, а квадратичная форма в круглых скобках положительно определенная. Мы предоставляем читателю подсчитать, что в этом случае $f(x) = -\frac{B}{C}x$, $\varphi(y) = -\frac{B}{A}y$, $r_{\xi,\eta} = -\frac{B}{\sqrt{AC}}$.

17. Характеристическая функция. *Характеристической функцией* случайной величины ξ называется функция от вещественного числового параметра u

$$\varphi_\xi(u) = \mathrm{M}\{e^{iu\xi}\} \qquad (-\infty < u < \infty),$$

На основании свойства 6 п. 14 нетрудно написать более подробное выражение характеристической функции в виде суммы или интеграла:

$$\varphi_\xi(u)=\sum_k P_k e^{iux_k} \quad \text{или} \quad \varphi_\xi(u)=\int e^{iux}p_\xi(x)\,dx. \tag{23}$$

Впервые характеристические функции в теории вероятностей систематически применил А. М. Ляпунов.

Вторая формула (23), с точностью до обозначений, совпадает с интегральным представлением Фурье (XVII.141). Значит, плотность $p_\xi(x)$ выражается через характеристическую функцию по формуле, аналогичной (XVII.138):

$$p_\xi(x)=\frac{1}{2\pi}\int \varphi_\xi(u)\,e^{-iux}\,du.$$

Укажем некоторые простые свойства характеристических функций. Так, из (23) ясно, что всегда $|\varphi_\xi(u)|\leqslant 1$, $\varphi_\xi(0)=1$. Если $\eta=C_1\xi+C_2$ $(C_1, C_2=\text{const})$, то

$$\varphi_\eta(u)=\mathbf{M}\left\{e^{iu(C_1\xi+C_2)}\right\}=\mathbf{M}\left\{e^{iC_2u}e^{iC_1u\xi}\right\}=e^{iC_2u}\varphi_\xi(C_1u).$$

Если $\zeta=\xi+\eta$, причем величины ξ и η независимые, то

$$\varphi_\zeta(u)=\mathbf{M}\left\{e^{iu(\xi+\eta)}\right\}=\mathbf{M}\left\{e^{iu\xi}e^{iu\eta}\right\}=\mathbf{M}\left\{e^{iu\xi}\right\}\mathbf{M}\left\{e^{iu\eta}\right\}=\varphi_\xi(u)\,\varphi_\eta(u).$$

Из первой формулы (23) и последнего свойства легко получаем характеристическую функцию для биномиально распределенной величины (п. 9): $\varphi_\xi(u)=(1-P+Pe^{iu})^n$. Отсюда с помощью перехода к пределу получаем для закона Пуассона $\varphi_\xi(u)=\exp(-\alpha+\alpha e^{iu})$. Для равномерного распределения (п. 10) $\varphi_\xi(u)=(e^{ibu}-e^{iau})/i(b-a)u$.

Для дальнейшего нам понадобится вычислить фурье-образ от функции $f(x)=\exp(-x^2)$, что, впрочем, имеет и самостоятельный интерес. По формуле (XVII.138) имеем (проверьте!)

$$\hat{f}(k)=\frac{1}{2\pi}\int \exp(-x^2-ikx)\,dx=\frac{1}{2\pi}\exp\left(-\frac{k^2}{4}\right)\int \exp\left[-\left(x+i\,\frac{k}{2}\right)^2\right]dx.$$

Однако последний интеграл, который мы обозначим через $I(k)$, на самом деле не зависит от k: действительно,

$$\frac{dI}{dk}=-\int \exp\left[-\left(x+i\,\frac{k}{2}\right)^2\right]2\left(x+i\,\frac{k}{2}\right)\frac{i}{2}\,dx=$$
$$=\frac{i}{2}\exp\left[-\left(x+i\,\frac{k}{2}\right)^2\right]\Bigg|_{x=-\infty}^{\infty}=0$$

(почему?). Значит, $I(k)=I(0)=\int \exp(-x^2)\,dx=\sqrt{\pi}$. Поэтому окончательно $\hat{f}(k)=\dfrac{1}{2\sqrt{\pi}}\exp\left(-\dfrac{k^2}{4}\right)$. Отсюда с помощью свойства 3 п. XVII.33 вытекает, что образом функции $\exp(-ax^2)$ $(a>0)$ служит $\dfrac{1}{2\sqrt{\pi a}}\exp\left(-\dfrac{k^2}{4a}\right)$.

Теперь легко вычислить характеристическую функцию для закона Гаусса (п. 10). Будем считать сначала, что $\alpha=0$. Так как формула (23) дает обратное преобразование Фурье, то надо дополнительно помножить результат на

2π, что даст

$$\varphi_\xi(u) = 2\pi \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{2} \sqrt{\frac{2\sigma^2}{\pi}} \exp\left(-\frac{u^2}{4} 2\sigma^2\right) = \exp\left(-\frac{\sigma^2 u^2}{2}\right).$$

Если теперь совершить сдвиг на α, то по свойству характеристических функций получаем

$$\varphi_\xi(u) = e^{i\alpha u} \exp\left(-\frac{\sigma^2 u^2}{2}\right) = \exp\left(i\alpha u - \frac{\sigma^2 u^2}{2}\right).$$

Отсюда, в частности, вытекает замечательное следствие. Пусть независимые величины ξ_1, ξ_2 распределены по закону Гаусса с параметрами α_1, σ_1 и α_2, σ_2 соответственно. Тогда для $\xi = \xi_1 + \xi_2$ получаем

$$\varphi_\xi(u) = \varphi_{\xi 1}(u)\, \varphi_{\xi 2}(u) = \exp\left(i\alpha_1 u - \frac{\sigma_1^2 u^2}{2}\right) \exp\left(i\alpha_2 u - \frac{\sigma_2^2 u^2}{2}\right) =$$
$$= \exp\left[i(\alpha_1 + \alpha_2)u - \frac{(\sigma_1^2 + \sigma_2^2)u^2}{2}\right].$$

Таким образом, получился опять закон Гаусса с параметрами $\alpha = \alpha_1 + \alpha_2$, $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Эта инвариантность закона Гаусса относительно сложения является одной из основных причин его распространенности. Для дискретных величин этим свойством обладает закон Пуассона.

§ 4. Применения нормального закона

18. Нормальный закон как предельный. Рассмотрим, как ведет себя биномиальный закон распределения (п. 9), если P остается неизменным, а $n \to \infty$. Если обозначить через $\xi^{(n)}$ случайную величину с таким законом распределения, то, как было уже показано, ее среднее значение $a_n = nP \to \infty$ и среднее квадратичное отклонение $\sigma_n = \sqrt{n}\sqrt{P(1-P)} \to \infty$. Это «размазывание» величины $\xi^{(n)}$ несколько затрудняет ее непосредственное исследование. Проще сначала сделать линейное преобразование величины $\xi^{(n)}$, после которого среднее значение стало бы равным нулю, а среднее квадратичное отклонение — единице. Это преобразование называется *нормировкой* величины $\xi^{(n)}$ и осуществляется по простой формуле

$$\eta^{(n)} = \frac{1}{\sigma_n}(\xi^{(n)} - a_n).$$

Замечательная *теорема Муавра — Лапласа* гласит, что *закон распределения нормированной таким образом величины при* $n \to \infty$ *стремится к закону Гаусса.* (П. Лаплáс, 1749—1827 — выдающийся французский астроном, математик и физик.)

В самом деле, в силу п. 17

$$\varphi_{\mu^{(n)}}(u) = e^{-i\frac{a_n}{\sigma_n}u}\left(1 - P + Pe^{i\frac{u}{\sigma_n}}\right)^n =$$
$$= \exp\left(-i\sqrt{\frac{nP}{1-P}}\,u\right)\left[1 - P + P\exp\left(\frac{iu}{\sqrt{nP(1-P)}}\right)\right]^n =$$
$$= \left\{\exp\left(-i\sqrt{\frac{P}{n(1-P)}}\,u\right)\left[1 - P + P\exp\left(\frac{iu}{\sqrt{nP(1-P)}}\right)\right]\right\}^n.$$

Разлагая выражение, стоящее в фигурных скобках, по степеням $\frac{1}{\sqrt{n}}$, получим

$$\varphi_{\eta(n)}(u)=\left\{\left(1-i\sqrt{\frac{P}{n(1-P)}}u-\frac{Pu^2}{2n(1-P)}+\ldots\right)\times\right.$$
$$\left.\times\left[1-P+P\left(1+\frac{iu}{\sqrt{nP(1-P)}}-\frac{u^2}{2nP(1-P)}+\ldots\right)\right]\right\}^n=$$
$$=\left\{1-\frac{u^2}{2n}+\ldots\right\}^n\xrightarrow[n\to\infty]{}\exp\left(-\frac{u^2}{2}\right)$$

(проверьте!). Но согласно п. 16 получилась как раз характеристическая функция закона Гаусса с $\alpha=0$, $\sigma=1$.

Мы уже отмечали в п. 14, что величина, распределенная по биномиальному закону, представляет собой сумму n независимых слагаемых, с одинаковым простейшим законом распределения. Оказывается, что тот же закон Гаусса получается в пределе при *любом* исходном законе распределения, кроме детерминированного.

В самом деле, допустим для простоты, что у исходного закона с характеристической функцией $\varphi_0(u)$ среднее значение равно нулю, а дисперсия — единице; так как проводится нормировка, то это ограничение несущественное. Тогда из формулы (23) вытекает, что $\varphi_0'(0)=0$, $\varphi_0''(0)=-1$, т. е. по формуле Тейлора $\varphi_0(u)=1-\frac{u^2}{2}+\ldots$ При обозначениях, аналогичных примененным выше, получим

$$\varphi_{\eta(n)}(u)=\left[\varphi_0\left(\frac{u}{\sqrt{n}}\right)\right]^n=\left[1-\frac{u^2}{2n}+\ldots\right]^n\xrightarrow[n\to\infty]{}\exp\left(-\frac{u^2}{2}\right).$$

Оказывается, что условие одинаковости законов распределения у слагаемых также необязательное. Так, А. М. Ляпунов доказал, что закон распределения нормированной суммы $\xi_1+\xi_2+\ldots+\xi_n$ независимых случайных слагаемых близок к закону Гаусса, если n велико, а отношение

$$\sum_{k=1}^{n}\mathbf{M}\,|\xi_k-a_k|^3:\left(\sum_{k=1}^{n}\mathbf{D}\xi_k\right)^{3/2}\qquad(a_k=\mathbf{M}\xi_k)$$

мало́. Это условие нарушается, если дисперсия небольшого числа слагаемых существенно превосходит дисперсию остальных слагаемых, которые поэтому после нормировки как бы отпадут; нарушается условие Ляпунова и в некоторых специальных случаях, например в случае, приводящем к закону Пуассона. (Проверьте!)

Если после нормировки получился закон Гаусса, то это значит, что и до нормировки был закон Гаусса, но уже с произвольными средним значением и дисперсией. Итак, *сумма многих независимых случайных слагаемых распределена по закону Гаусса, независимо от того, как были распределены слагаемые*; исключение составляют

лишь случаи, указанные в предыдущем абзаце. Это и есть важнейшая причина огромного значения закона Гаусса.

В частности, обычно принимается, что ошибки измерения подчиняются закону Гаусса. В самом деле, такая ошибка обычно складывается из многих независимых трудно контролируемых мелких ошибок, что и приводит к этому допущению.

19. Доверительные оценки средних. Вернемся к задаче о бросании монеты, рассмотренной в п. 1. Ясно, что если из 1000 бросаний герб выпадет 200 раз, то что-то не в порядке. А если 400, 450 раз? Другими словами, может ли после 1000 бросаний относительная частота выпадания герба равняться 0,4 или 0,45? Теперь мы в состоянии ответить на этот вопрос.

Рассмотрим более общую задачу. Пусть рассматривается некоторая случайная величина ξ (в предыдущем примере это было число выпаданий герба при одном бросании). Пусть проведено n испытаний, в результате которых она приняла значения $x_1, x_2, \ldots, x_n$. Обозначим среднее арифметическое этих значений через

$$x^{(n)} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n). \tag{24}$$

На основе п. 13 мы можем утверждать, что при $n \to \infty$ будет $x^n \to \bar{\xi}$; этот *закон больших чисел* при нашем изложении мы приняли за основу определения среднего $\bar{\xi}$. Но с какой скоростью $x^{(n)} \xrightarrow{} \bar{\xi}$?

Рассмотрим случайную величину

$$\xi^{(n)} = \frac{1}{n}(\xi_1 + \xi_2 + \ldots + \xi_n),$$

где все слагаемые независимые и распределены по тому же закону, что ξ. Тогда (24) — это одно из значений величины $\xi^{(n)}$. Однако в силу п. 18 при большом n можно принять, что величина $\xi^{(n)}$ распределена по закону Гаусса. Притом

$$\overline{\xi^{(n)}} = \frac{1}{n}(\bar{\xi}_1 + \bar{\xi}_2 + \ldots + \bar{\xi}_n) = \bar{\xi},$$

$$\mathbf{D}\xi^{(n)} = \frac{1}{n^2} n\mathbf{D}\xi = \frac{1}{n}\mathbf{D}\xi = \frac{1}{n}\sigma_\xi^2,$$

где σ_ξ — среднее квадратичное уклонение величины ξ. Поэтому в силу формулы (14) можно принять, что

$$\mathbf{P}\{a \leqslant \xi^{(n)} \leqslant b\} = \frac{1}{2}\left[\Phi\left(\sqrt{n}\,\frac{b-\bar{\xi}}{\sigma_\xi}\right) - \Phi\left(\sqrt{n}\,\frac{a-\bar{\xi}}{\sigma_\xi}\right)\right].$$

Ограничимся для простоты симметричными интервалами вида $|\xi^{(n)} - \bar{\xi}| \leqslant \delta$:

$$\mathbf{P}\{|\xi^{(n)} - \bar{\xi}| \leqslant \delta\} = \Phi\left(\frac{\delta\sqrt{n}}{\sigma_\xi}\right). \tag{25}$$

Эта формула и решает поставленную задачу. Практически ею можно пользоваться при любых n. Для ориентировки приведем самую грубую таблицу функции $\Phi(t)$:

t	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1
$\Phi(t)$	0,00	0,08	0,16	0,24	0,31	0,38	0,45	0,52	0,58	0,63	0,68	0,73

t	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	2,1	2,2	2,3	2,4
$\Phi(t)$	0,77	0,81	0,84	0,87	0,89	0,91	0,93	0,94	0,95	0,96	0,97	0,98	0,98

t	2,6	3,0	3,3	3,9	4,4	4,9	5,3	∞
$\Phi(t)$	0,99	0,997	0,999	0,9999	0,99999	6 девяток	7 девяток	1

Как же быть с бросанием монеты? Условимся начать сомневаться, если произойдет событие, вероятность которого меньше 0,01. Тогда, выбрав δ таким, чтобы правая часть (25) равнялась 0,99, мы полностью исчерпаем область значений, не вызывающих сомнения. Отсюда $\frac{\delta\sqrt{n}}{\sigma_\xi}=2{,}6$, и так как в нашем примере $n=1000$, $\sigma_\xi=\frac{1}{2}$, мы получаем, что $\delta=0{,}041$. Итак, мы получаем *доверительные оценки* $459\leqslant M\leqslant 541$ для числа M выпаданий герба, т. е. оценки, которые мы гарантируем, игнорируя события, маловероятные относительно выбранного критерия (п. 7).

20. Обработка эмпирических данных. В п. 19 мы считали, что закон распределения случайной величины ξ задан. Однако на практике чаще встречаются задачи, в которых этот закон не задан, а числовые характеристики величины ξ требуется найти, исходя из результатов испытаний. Такая задача возникает, если требуется определить среднее значение какой-либо числовой характеристики (скажем, прочности, долговечности и т. п.) изделия в некоторой партии (*генеральной совокупности*) изделий с помощью проверки этой характеристики у выбранных наугад контрольных изделий. Та же задача возникает при повторном измерении какой-либо неизвестной физической величины, так как из-за возможных ошибок измерения его результат является случайной величиной, среднее значение которой при отсутствии систематических ошибок измерения равно точному значению измеряемой величины, и т. д.

Пусть произведено n испытаний и величина ξ приняла значения $x_1, x_2, \ldots, x_n$. Тогда из предыдущего вытекает, что за приближенное значение для $\overline{\xi}$ следует принять среднее арифметическое (24) из полученных результатов. Кроме того, оказывается, что за при-

ближенное значение для среднего квадратичного уклонения σ_ξ можно принять

$$\sigma_\xi \approx \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - x^{(n)})^2}. \tag{26}$$

Для доказательства этого воспользуемся обозначениями п. 19 и обозначим дополнительно $\zeta_n = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \xi^{(n)})^2$. Подсчитаем среднее значение (проверьте!)

$$\mathbf{M}\{\zeta_n\} = \frac{1}{n-1}\mathbf{M}\left\{\sum_{i=1}^{n}\xi_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\xi_i\right)^2\right\}.$$

Представляя каждое ξ_i в виде $(\xi_i - \bar{\xi}) + \bar{\xi}$ и производя вычисления, получаем

$$\frac{1}{n-1}\left[n\mathbf{D}\xi + n\bar{\xi}^2 - \frac{1}{n}(n\mathbf{D}\xi + n^2\bar{\xi}^2)\right] = \mathbf{D}\xi = \sigma_\xi^2.$$

Кроме того, подсчет $\mathbf{D}\zeta_n$, который мы предоставляем читателю, показывает, что $\mathbf{D}\zeta_n \to 0$ при $n \to \infty$. Однако подкоренное выражение в (26) — это значение случайной величины ζ_n. Отсюда и следует (26).

Теперь можно, задавшись критерием маловероятности (п. 7), на основании формулы (25) получить доверительные границы для $\bar{\xi}$. Например, будем игнорировать события, вероятность которых менее 0,003. Тогда из (25) следует, что можно принять $\frac{\delta\sqrt{n}}{\sigma_\xi} = 3$, и мы получаем довольно распространенную доверительную оценку для среднего значения

$$x^{(n)} - \frac{3\sigma_\xi}{\sqrt{n}} \leqslant \bar{\xi} \leqslant x^{(n)} + \frac{3\sigma_\xi}{\sqrt{n}}, \tag{27}$$

выполняющуюся с вероятностью 0,997. В этой формуле σ_ξ можно подставить из (26).

Обратим внимание на одно следствие из оценки (27). Пусть для одной и той же величины ξ проведено две серии из n испытаний, что привело к средним значениям $x^{(n)}$ и $\tilde{x}^{(n)}$. Тогда с вероятностью $0{,}997^2 = 0{,}994$ для обоих этих значений выполнены оценки (27), откуда $|x^{(n)} - \tilde{x}^{(n)}| \leqslant \frac{6\sigma_\xi}{\sqrt{n}}$. Значит, для больших n эмпирические средние $x^{(n)}$ с большой вероятностью принимают вполне определенное значение в любой серии испытаний. Впрочем, этот результат, хотя и без количественной оценки, вытекает из самого определения среднего значения.

Если рассматривается пара случайных величин ξ, η, находящихся в линейной корреляционной взаимосвязи (п. 16), то после испытаний получается система пар значений $(x_1;\ y_1)$, $(x_2;\ y_2)$, $\ldots$, $(x_n;\ y_n)$. Можно поставить задачу о приближенном построении функции регрессии η на ξ на основе этих данных. Оказывается, что решение этой задачи приводит как раз к методу наименьших квадратов, описанному в п. XII.8.

Теория вероятностей, зародившаяся в XVII в. в связи с анализом азартных игр, особенно развилась в XIX и XX веках и в настоящее время является весьма содержательной дисциплиной, богатой многочисленными приложениями. Большой вклад в нее внесли П. Л. Чебышев, А. М. Ляпунов, А. А. Марков, С. Н. Бернштейн, А. Н. Колмогоров, Ю. В. Линник и многие другие отечественные и зарубежные ученые. С теорией вероятностей непосредственно связана *математическая статистика*, изучающая методы обработки статистических данных. Имеется много курсов теории вероятностей, математической статистики и их приложений; из более кратких отметим [1], [9], [11], [40], [44].

О СОВРЕМЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКЕ

Простейшие вычислительные средства — логарифмическая линейка, конторские счеты, арифмометр, таблицы уже давно внедрены в практику и широко известны. Однако для потребностей современной науки, техники, экономики эти средства недостаточны. Многие важные задачи, в принципе разрешимые, требуют очень большой вычислительной работы, так что их решение с помощью указанных средств невозможно получить в практически приемлемые сроки. Для получения решения часто приходилось не принимать во внимание многие существенные факторы, что приводило к количественным, а иногда и к принципиальным ошибкам.

Поиски более эффективных вычислительных средств и достижения современной техники привели к подлинному перевороту в этой области, который произошел в последнее двадцатилетие и проявился в создании быстродействующих электронных вычислительных машин. Интенсивное внедрение этих машин в практику позволило успешно разрешить ряд важных задач и значительно увеличить количество областей человеческой деятельности, в которых применение математики приводит к полезным результатам. Несомненно, что дальнейшее развитие и, особенно, расширение круга применений современной вычислительной техники приведут уже при жизни нашего поколения к коренной перестройке организации научных исследований, технических работ, экономических расчетов, управления, обслуживания и т. д.

§ 1. Два основных типа вычислительных машин

Имеются два основных способа представления математических величин, над которыми производятся вычисления. По одному способу такие величины *непосредственно* изображаются теми или иными физическими величинами — длинами, углами, напряжениями и т. д., после чего *подбирается физическая схема, в которой физические величины преобразуются по тому же закону, по которому должны преобразовываться математические величины*. Вычислительные машины, основанные на таком принципе, называются *аналоговыми* или *моделирующими*; простейшими аналоговыми приборами являются логариф-

мическая линейка и планиметр (п. XIV. 12). По второму способу с помощью какой-либо конструкции изображается *цифровая запись рассматриваемых математических величин, а действия над ними заменяются арифметическими действиями над цифрами.* Вычислительные машины, основанные на этом принципе, называются *цифровыми*; к ним относятся, в частности, конторские счеты и арифмометр. Наибольшие достижения, о которых говорилось выше, относятся к цифровым машинам.

1. Моделирующие вычислительные машины. Рассмотрим сначала простейший пример. Пусть требуется найти сумму s двух заданных величин x и y:

$$s = x + y. \tag{1}$$

Эту задачу можно промоделировать на механической схеме (рис. 346, *а*), приняв x, y и s за длины, либо на электрической схеме с переменными сопротивлениями (рис. 346, *б*), приняв x, y и s за токи, а также на многих других схемах. Уже на этом примере хорошо видны характерные черты моделирующих вычислительных машин.

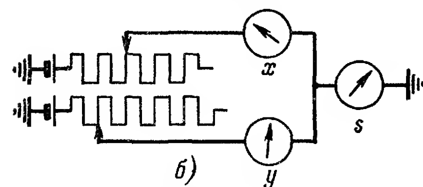


Рис. 346

Прежде всего, *входные параметры* x и y могут непрерывно изменяться в определенных диапазонах; поэтому данный класс машин называют также вычислительными машинами *непрерывного действия.* Конечно, можно себе представить, что на рис. 346, *б* сопротивления меняются дискретно, с помощью «магазина сопротивлений»; но эта дискретность была бы конструктивной, непринципиальной, тогда как цифровые машины принципиально являются машинами *дискретного действия.*

Далее, ясно, что точность ввода параметров и получения результата невелика; и действительно, она обычно бывает порядка процентов, в лучшем случае десятых долей процентов. Это, конечно, ограничивает возможности моделирования громоздких вычислений. Кроме того, моделирующие машины обычно являются сравнительно узко специализированными, т. е. приспособленными лишь для решения задач определенного узкого класса. Так, приспособления, изображенные на рис. 346, предназначены лишь для выполнения действия (1) и непосредственно связанных с ним действий, например вычитания. В то же время, если такие специализированные задачи приходится решать часто и особой точности не требуется, то применение моделирующих машин и приборов оказывается очень эффективным. Например, довольно широко применяются *электрические интеграторы,*

предназначенные для решения систем обыкновенных дифференциальных уравнений.

На разобранном примере хорошо видно также, что *одну и ту же* взаимосвязь величин можно осуществить на *различных* физических схемах. Это имеет место и во многих гораздо более важных и сложных примерах и дает основание для *моделирования* физических процессов: чтобы в некоторой физической схеме найти какую-либо величину S, которую трудно замерить или подсчитать непосредственно, создают новую схему, иной физической природы, но в которой между участвующими величинами имеет место такая же функциональная взаимосвязь; после этого в новой схеме замеряют величину, соответствующую S. При этом, если математическая эквивалентность двух физических схем установлена, то математическое решение задачи оказывается излишним, так что никаких вычислений, если они не нужны для чего-либо иного, можно не производить. В последние годы такое *аналоговое* решение задач на основе электромеханических, оптико механических, электродиффузионных и т. д. аналогий получило широкое развитие.

Часто моделирующая вычислительная машина представляет собой агрегат, состоящий из нескольких *блоков* (элементов), каждый из которых предназначен для выполнения какого-либо одного действия. Рассмотрим, например, наиболее распространенный сейчас способ представления величин с помощью электрических напряжений. Суммирующий блок, схематически изображенный на рис. 347, *а*, обладает двумя входами и одним выходом, причем если к входным клеммам подключить постоянные или зависящие от времени t напряжения x и y, то на выходной клемме напряжение будет равняться $z = x + y$. При этом, как именно происходит суммирование, нас в данном случае не интересует. Аналогичный вид имеет множительный блок (рис. 347, *б*), блок функционального преобразования (рис. 347, *в*) при вполне определенной функции f (скажем, синусный блок при $f(x) \equiv \sin x$ и т. п.); блок усиления (рис. 347, *г*), в котором коэффициент усиления k можно менять, и т. д. При решении дифференциальных уравнений применяется дифференцирующий блок (рис. 347, *д*), который при входных, зависящих от времени напряжениях $x(t)$ и $y(t)$ вырабатывает выходное напряжение dy/dx, также, вообще говоря, зависящее от времени. Применяются и другие блоки. Собирая эти блоки в различные агрегаты, мы получаем возможность решать различные уравнения

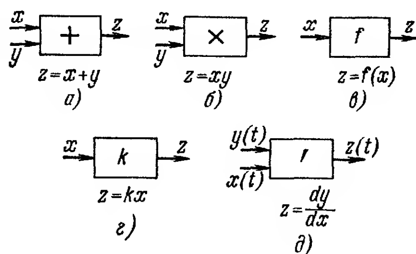


Рис. 347.

Покажем, например, схему для решения системы уравнений

$$\left.\begin{aligned} ax+by&=c,\\ dx+f(y)&=0.\end{aligned}\right\}$$

Для этого ее удобно переписать в виде

$$\left.\begin{aligned} y&=\frac{c}{b}-\frac{a}{b}x,\\ x&=-\frac{1}{d}f(y).\end{aligned}\right\}$$

Соответствующая схема изображена на рис. 348, причем кружком обозначен источник напряжения, с одной стороны заземленный, и на каждом канале указано передаваемое напряжение. Величины $\frac{1}{b}$ и $\frac{1}{d}$ не надо подсчитывать, так как шкалу коэффициента усиления можно разградуировать в обратных значениях.

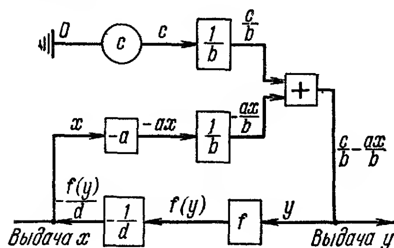


Рис. 348.

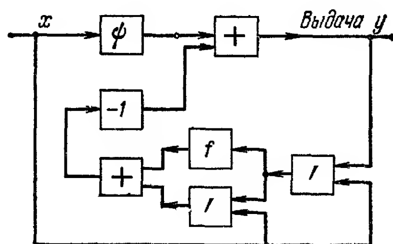


Рис. 349.

Хорошо видно, что на данной схеме удобно изучать влияние коэффициентов на решение, а при помощи видоизменения схемы можно решать и другие аналогичные системы. Однако надо иметь в виду, что если задача имеет несколько решений, то схема может дать не то решение, которое нужно.

Покажем еще схему для решения дифференциального уравнения, например, вида

$$y''+f(y')+y=\psi(x) \qquad (y=y(x)),$$

для чего перепишем его так:

$$y=\psi(x)-y''-f(y').$$

Соответствующая схема показана на рис. 349 (проверьте ее правильность!). Здесь y уже является переменной величиной, изменение которой следует отнести к изменению независимой переменной x.

Часто в качестве независимой переменной берется время t. Тогда дифференцирующий блок имеет только один вход, а в разобранном втором примере ввод x не требуется; просто на клемму, обозначенную на рис. 349 точкой, надо подавать напряжение $\psi(t)$, что можно делать без предварительных вычислений, прямо с графика или какого-либо иного датчика. Это особенно удобно, если и по физическому смыслу задачи независимой переменной является t, т. е. задача

решается в *реальном масштабе времени*. Тогда решение задачи можно непосредственно, минуя человека, передавать в какую-либо иную схему для использования; на этом принципе основаны многие приспособления для автоматического регулирования.

Применение реального масштаба времени дает также возможность при испытании сложных конструкций заменять отдельные дорогостоящие агрегаты вычислительными машинами. Например, автопилот можно испытывать не в полете, что дорого и опасно, а на стенде, заменив самолет моделирующей машиной, которая по своей реакции на работу автопилота полностью тождественна самолету.

Основы теории математических машин непрерывного действия изложены, например, в книгах [30], [52].

2. Цифровые вычислительные машины. Первые приспособления типа конторских счетов были созданы в Китае 2—3 тысячи лет до н. э., а первую суммирующую машину сконструировал в 1641 г. выдающийся французский математик и физик, философ Б. Паскáль (1623—1662). Результатом усовершенствования машины Паскаля явились широко распространенные сейчас *малые счетные машины* — суммирующие машины, выполняющие действия сложения и вычитания, и полуавтоматические и автоматические арифмометры, производящие все четыре арифметических действия. Все эти чрезвычайно полезные машины, конечно, не могут действовать очень быстро как из-за невысокой скорости самого вычисления, так и из-за того, что входные данные для каждого арифметического действия вводятся человеком-вычислителем.

Потребности обработки большого числа однородных данных, возникающие в статистике, бухгалтерском учете и финансово-банковском деле, привели к созданию в начале XX в. *счетно-аналитических* (счетно-перфорационных) *машин* — перфорáторов, сортировок, табуля́торов и т. д. В такие машины входные параметры поступают пробитыми на перфокартах (см. п. 4); с помощью «прощупывания» этих карт системой электрических контактов машина их сортирует по заданной программе, производит суммирование параметров и накопление итогов, а также другие простые действия. Для более сложных действий эти полезные машины приспособлены мало.

Как было сказано, переворот в этой области наступил в связи с созданием быстродействующих автоматических универсальных цифровых машин. Хотя идея конструкции таких машин была высказана еще в XIX в., но лишь достижения современной электроники сделали возможной ее реализацию. Первая быстродействующая электронная цифровая машина *ЭНИАК* была построена в США в 1943 г., а в 1946 г. американский математик Дж. Нéйман (1903—1957) сформулировал основные идеи и принципы построения таких машин. В СССР первые машины этого типа были пущены в 1952 г.; сейчас у нас имеется большое число машин различной мощности и разного назначения (п. 7).

Структурная схема (блок-схема) автоматической цифровой вычислительной машины, т. е. основные ее блоки и их взаимосвязь, показана на рис. 350 В *запоминающем устройстве* (ЗУ) имеется определенное количество, например 512, *ячеек*, каждая из которых служит для записи одного числа или одной «команды» (предписания). Запись команды, как мы увидим в п. 5, по внешней форме ничем не отличается от записи числа. Некоторые из этих ячеек заполняются от устройства ввода, тогда как другие заполняются по мере работы, причем в процессе работы содержимое ячеек может многократно заменяться, а часть ячеек не используется. Руководствуясь записанными командами, *устройство управления* (УУ) направляет числа (или команды) из ЗУ в *арифметическое устройство* (АУ), которое их преобразует в соответствии с этими командами и возвращает в ЗУ. По мере получения требуемых результатов они по сигналам УУ автоматически печатаются, а после окончания вычисления УУ останавливает машину. (Подробнее эта последовательность действий будет описана в § 2.)

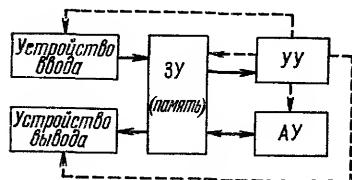


Рис 350.

Каналы информации обозначены сплошными стрелками, каналы управления обозначены пунктирными стрелками

На *универсальной* автоматической цифровой машине можно решить любую математическую задачу, для которой указан *алгорйтм*, т. е. четкое предписание, определяющее вычислительный процесс, которое, исходя из входных данных, обязательно приводит к требуемому результату. Однако и нематематические задачи, для которых можно указать аналогичный алгоритм решения, разрешимы на таких машинах. Например, цифровая машина может управлять металлорежущим станком при обработке детали как угодно сложного профиля; при этом вводные данные характеризуют профиль, а результаты вычисления преобразуются в сигналы устройству управления станком. Аналогично работает устройство, управляющее полетом самолета, начиная от взлета и кончая посадкой в заданном месте. Цифровая машина может управлять производственным процессом, что особенно важно во вредных для человека отраслях производства, и дает возможность полностью автоматизировать процесс. В случае отклонения от первоначально заданной программы машина может принимать наивыгоднейшие решения, сравнивая возможные варианты, а также контролировать результат. Конечно, при этом информация о действительном состоянии процесса должна вводиться в машину автоматически, с помощью специальных устройств. При разработке инженерной конструкции машина может просчитывать десятки и сотни вариантов, чтобы выбрать из них наилучший по какому-либо заданному признаку. Аналогичные применения получили цифровые машины для

прогноза погоды, для планирования перевозок, размещения предприятий и т. д. Впрочем, во всех этих применениях более эффективными обычно оказываются не универсальные, а *специализированные* цифровые машины, приспособленные для решения определенного круга задач.

Это открывает широкие возможности автоматизации многих видов умственного труда человека; человек вырабатывает алгоритм, а вычислительная машина его реализует, иногда даже вырабатывая новые алгоритмы в процессе этой реализации.

§ 2. Понятие о программировании

3. Системы счисления. Десятичная система счисления, которую мы изучаем в школе и постоянно применяем в жизни, неудобна для использования в работе электронных цифровых машин. Здесь чаще всего применяется *двоичная* система счисления, в которой имеется всего лишь две цифры, 0 и 1 (тогда как в десятичной системе имеется десять цифр), и все числа записываются в виде комбинаций этих цифр. Для этого принимают, что записи 10, 100 и т. д. означают не степени десятки, а степени двойки, т. е. числа два, четыре и т. д.

Покажем перевод первых натуральных чисел из десятичной системы в двоичную.

Десятичная запись	1	2	3	4	5	6	7	8	9	10	11	12	13	и т. д.
Двоичная запись	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	и т. д

Любое целое число можно записать в двоичной системе счисления; для этого из данного числа надо выделять степени двойки, начиная со старшей. Например,

$$1967 = 1\cdot 1024 + 1\cdot 512 + 1\cdot 256 + 1\cdot 128 + 0\cdot 64 + 1\cdot 32 + 0\cdot 16 + \\ + 1\cdot 8 + 1\cdot 4 + 1\cdot 2 + 1\cdot 1,$$

т. е. двоичная запись этого числа такова: 11 110 101 111.

Аналогично составляются двоичные дроби, взамен десятичных. Например, двоичное число 10,1011 означает $2+\frac{1}{2}+\frac{1}{8}+\frac{1}{16}=\frac{43}{16}$. Любое нецелое число можно записать в виде конечной или бесконечной двоичной дроби; естественно, что в реальных вычислениях дроби округляются.

Таблицы сложения и умножения в двоичной системе крайне просты:

$$0+0=0,\ 1+0=0+1=1,\ 1+1=10,$$

$$0\cdot 0=0,\ 0\cdot 1=1\cdot 0=0,\ 1\cdot 1=1.$$

Пользуясь этими таблицами, возможно производить арифметические действия над числами, записанными в двоичной системе, совершенно аналогично тому, как это делается в школе для десятичной системы.

Основной недостаток двоичной системы состоит в том, что двоичная запись сравнительно небольших целых чисел занимает очень много места.

Поэтому для вспомогательных записей применяются и другие системы, в особенности *восьмеричная*. В ней имеется восемь цифр: 0, 1, 2, 3, 4, 5, 6, 7; число «восемь» записывается в виде 10, «девять» — в виде 11, ..., «шестьдесят четыре» — в виде 100 и т. д. Легко переходить от восьмеричной записи к десятичной и обратно; например, восьмеричное число 571 в десятичной системе запишется в виде

$$5\cdot 8^2+7\cdot 8+1=377.$$

Применение восьмеричной системы объясняется тем, что длина записи чисел в ней не намного превышает длину записи в десятичной системе и в то же время от восьмеричной записи числа крайне просто переходить к двоичной записи. Например, числа 5, 7 и 1 имеют двоичную запись 101, 111 и 001 и потому восьмеричное число 571 имеет двоичную запись 101 111 001, так как переходу в следующий разряд в восьмеричной системе отвечает умножение на 1000 в двоичной системе. В частности, нумерация ячеек производится по восьмеричной системе: так, если ячеек 512, то они получают номера от 0 до 777 (проверьте!).

Для ввода чисел в машину применяется своеобразная *двоично-десятичная* система. В ней первые десять (начиная от нуля) чисел записываются, как в двоичной системе:

Десятичная запись	0	1	2	3	4	5	6	7	8	9
Двоично-десятичная запись	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001

Для записи дальнейших натуральных чисел взамен каждого десятичного знака подставляется соответствующая четверка цифр из 0 и 1. Например, числа 63 и 125 в двоично-десятичной записи имеют вид

0110 0011 и 0001 0010 0101.

Аналогично преобразуются десятичные дроби в двоично-десятичные дроби.

Двоично-десятичная запись еще длиннее, чем двоичная. Однако ее удобство в том, что от нее крайне просто переходить к десятичной и обратно.

4. Запись чисел в машине. В электронной цифровой машине числа записываются по двоичной системе в ЗУ (п. 2) в ячейках одинаковой длины, т. е. с одинаковым числом *разрядов*, каждый из которых содержит 0 или 1. Для каждой конструкции машин принят один из двух способов записи:

1. При записи с *фиксированной запятой* первый разряд ячейки отводится под знак числа: + изображается в виде 0, а — в виде 1; дальнейшие разряды ячейки отводятся под последовательные дробные (после запятой) разряды двоичного разложения этого числа. Например, если в ячейке 30 разрядов, то запись

1	0	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(2)

означает число $-0{,}0010111001$ в двоичной записи, т. е.

$$-\left(\frac{1}{8}+\frac{1}{32}+\frac{1}{64}+\frac{1}{128}+\frac{1}{1024}\right)=-\frac{185}{1024}=-0{,}1806640625$$

в десятичной записи.

При данной длине ячейки возможно записать любое число от $-(1-2^{-29})$ до $+(1-2^{-29})$ с шагом 2^{-29} (почему?). Поэтому при вычислениях во все участвующие величины вводится, если это требуется, масштабный множитель с тем, чтобы все они попали на интервал от -1 до $+1$.

Отметим, что применяются и другие способы записи отрицательных чисел; эти способы мы не будем рассматривать.

2. При записи *с плавающей запятой*, впервые примененной в советских машинах, диапазон изображаемых чисел значительно расширяется за счет дополнительного умножения на степень двойки. При этом под показатель степени, включая его знак, отводится некоторое вполне определенное число разрядов, например, в конце ячейки. Если этих разрядов шесть, то при общей длине ячейки в 30 разрядов наибольшее число, которое можно записать в ячейке, имеет такой вид:

знак числа ↓ — дробные разряды числа — знак показателя — разряды показателя

0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1

(3)

и равно $(1-2^{-23})\,2^{31}=2^{31}-2^{8}\approx 2^{31}$, тогда как наименьшее положительное число, которое можно записать по данному способу,

имеет вид

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1

(4)

и равно $2^{-23}2^{-31}=2^{-54}$ (а не 2^{-29}, как в первом способе).

Бывает так, что имеющегося числа разрядов недостаточно для проведения вычислений с требуемой точностью. Тогда можно программой работы предусмотреть, чтобы каждые *две* соседние ячейки — первая и вторая, третья и четвертая и т. д. — представляли разложение *одного* числа (*удвоение ячеек*).

Фактическая «запись» цифр для разных конструкций машин производится различными способами. Для ввода чисел во вводное устройство машины (п. 2) они пробиваются на специальной ленте — *перфоленте* или на наборе специальных карточек — *перфокарт*. В обоих случаях каждой ячейке отвечает участок определенной длины на ленте или строка на карточке, причем длина строки соответствует

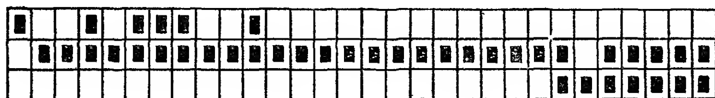


Рис. 351.

числу разрядов в ячейке, пробитое отверстие соответствует цифре 1, а непробитое — цифре 0. Например, если бы числа (2), (3) и (4) следовали одно за другим во вводной программе, то соответствующая перфокарта имела бы участок вида, изображенного на рис. 351 (проверьте!). Эти отверстия пробиваются перед запуском машины на специальных *перфораторах*, не связанных с машиной.

Часто при составлении программы программист записывает команды (п. 5) по восьмеричной системе, а входные числа — по десятичной. Тогда после нажатия клавишей перфоратора (он внешне напоминает кассовый аппарат в магазинах) в соответствии с этой записью на перфокартах команды оказываются пробитыми по двоичной системе, а входные числа — по двоично-десятичной системе. Переход к двоичной системе предусматривается программистом и осуществляется внутри машины в самом начале ее работы по специальной стандартной программе

Запись на перфокартах удобнее, чем запись на перфоленте, так как программа часто бывает очень обширной и записывается на колоде перфокарт, а заменить несколько карт из колоды, если обнаружена ошибка в программе или если входные данные изменились, гораздо проще, чем проводить изменения в ленте.

Для хранения (и преобразования по мере вычислений) чисел в ЗУ внутри машин каждая ячейка ЗУ реализуется в виде набора *элементов*, каждый из которых может находиться в одном из двух

состояний; число элементов в наборе равно числу разрядов в ячейке, а каждое из состояний элемента отвечает значению 0 и 1 в соответствующем разряде. В различных конструкциях машин применяются различные типы таких элементов, однако от них всегда требуется *безынерционность*, т. е. мгновенный переход из одного состояния в другое, и *устойчивость*, т. е. неопределенно долгое сохранение состояния при отсутствии переходного сигнала. Элемент может иметь вид *триггера*, состоящего из двух триодов, у которых катоды соединены вместе, а аноды и сетки — крест-на́крест, так что всегда либо один триод заперт, либо другой, причем переход из одного состояния в другое осуществляется с помощью импульса напряжения. Элементами могут служить заряженные и незаряженные участки на диэлектрической пластинке, причем переход из одного состояния в другое осуществляется с помощью электронно-лучевой трубки, подобной трубке телевизора. Элементами могут служить намагниченные и ненамагниченные участки на быстровращающемся магнитном барабане, действующем по принципу магнитофона. Применяются также магнитные ферритовые сердечники, которые могут менять направление магнитного поля при прохождении тока по обмотке, и т. д.

Передача числа из ячейки ЗУ в АУ и обратно может производиться либо по одному каналу с помощью последовательности импульсов для машин *последовательного действия*, либо по системе каналов, число которых может доходить до числа разрядов в ячейке, для машин *параллельного действия*. Второй способ быстрее, но усложняет конструкцию машины.

Чем больше ячеек в ЗУ (реальные количества ячеек будут указаны в п. 7), тем больше данных можно ввести в машину. Поэтому для расширения объема памяти многим машинам придается *внешнее* ЗУ, обычно оформляемое в виде магнитной ленты, с которой специальное устройство по заранее предусмотренным командам может вводить записанные там числа во внутреннее, *оперативное* ЗУ. Число ячеек на магнитной ленте может доходить до миллионов, однако считывание чисел с ленты и запись их обратно на ленту осуществляется не так быстро, как внутри оперативного ЗУ, так как требуется определенное время на перемещение ленты.

Результаты вычислений по стандартной программе переводятся в десятичную систему и печатаются на бумажной ленте или остаются в двоичной записи во внешнем ЗУ для использования в каких-либо дальнейших вычислениях.

Если машина перерабатывает не числа, а какую-либо иную информацию, например словесную, то эта информация должна быть предварительно *закодирована* (зашифрована) с помощью двоичных чисел. Тогда можно предусмотреть, чтобы после окончания действий специальное устройство раскодировало результат, т. е. перевело его в форму, естественную для данного вида информации. Это делается, например, при машинном переводе с одного языка на другой.

5. Команды. Машина производит действия над числами, хранящимися в ЗУ, только по *командам*, которые также вводятся перед началом работы в ЗУ и по внешней записи не отличаются от чисел. Применяются *одноадресные*, *двухадресные* и *трехадресные* команды, в зависимости от конструкции машин. Рассмотрим наиболее простые для программирования трехадресные команды.

Трехадресная команда представляет собой ячейку, мысленно разделенную на четыре части вполне определенной длины, например, в 3, 9, 9 и 9 разрядов. Эти части отведены под:

1) условную запись (*код*) действия, которое надо произвести;

2) *адрес* (номер ячейки) первого из чисел, участвующих в действии;

3) адрес второго из чисел, участвующих в действии;

4) адрес, по которому надо отправить результат действия.

При этом все ячейки нумеруются в естественном порядке. Например, если в ЗУ имеется 512 ячеек, то девяти двоичных разрядов как раз достаточно для указания адреса (почему?). Пусть кодом операции сложения служит число 1. Тогда если требуется сложить числа, находящиеся в 417-й и 73-й ячейках, а результат поместить в 646-ю ячейку, то при принятом выше разделении ячейки соответствующая команда должна иметь вид (проверьте, имея в виду, что номера восьмеричные!)

$$\text{части: } \underbrace{001}_{1\text{-я}}\ \underbrace{100001111}_{2\text{-я}}\ \underbrace{000111011}_{3\text{-я}}\ \underbrace{110100110}_{4\text{-я}} \qquad (5)$$

После выполнения этой команды во всех ячейках ЗУ, включая 417-ю и 73-ю, останется то же, что и было, за исключением 646-й, в которой взамен того, что в ней было, появится сумма чисел, стоящих по адресам 417 и 73.

Команда (5) пробивается на перфоленте или перфокарте, а затем хранится в одной из ячеек ЗУ в точности так же, как и числа. Различие проявляется лишь во время работы машины, так как если программа работы составлена правильно, то в УУ в качестве команд поступают сигналы лишь от тех ячеек, которые и были предусмотрены как команды.

Одноадресная команда представляет собой ячейку, разделенную на две части, в которых указывается код операции и номер ячейки. Чтобы произвести рассмотренное выше действие, требуются три одноадресные команды: по одной число из 417-й ячейки засылается в *сумма́тор*, по другой к содержимому сумматора прибавляется содержимое 73-й ячейки, а по третьей результат сложения отправляется в 646-ю ячейку. Двухадресная команда делится на три части, в которых указываются код операции и адреса чисел, над которыми производятся действия; при этом результат действия помещается по второму адресу.

Впредь мы будем для простоты рассматривать только трехадресные команды и записывать вместо кода операции знак действия, например, знак $+$ вместо кода сложения, а вместо восьмеричных номеров ячеек — их десятичные номера.

В советских машинах принят *естественный* (последовательный) порядок выполнения команд. В самом начале работы машины в ЗУ с перфокарт или перфоленты вводятся *программа* работы, т. е. определенная последовательность команд, и *исходные данные*, т. е. числовой массив, над которым производятся действия. Затем УУ воспринимает содержимое первой ячейки в качестве команды, и в соответствии с этой командой АУ выполняет предписанные действия. Далее УУ воспринимает содержимое второй ячейки в качестве команды и т. д., за исключением особых команд «передачи управления», о которых мы скажем в п. 6 и после выполнения которых УУ переходит не к очередной ячейке, а к ячейке, номер которой указан в команде. Некоторые команды предписывают печатание содержимого какой-либо ячейки, но и после выполнения такой команды УУ переходит к следующей ячейке. (Правда, если печатать приходится часто, то темп работы замедляется, так как печатание производится медленнее, чем арифметические операции.) И так продолжается до тех пор, пока УУ не доходит до команды, означающей *останов* машины, после чего работа прекращается. Машина останавливается (*аварийный останов*) и в том случае, если действие, предусмотренное программой, невыполнимо, например, если число, из которого предписано извлечь квадратный корень, отрицательно. То же получается при *переполнении ячеек*, т. е. если по ходу вычислений получится столь большое число, что оно не может быть записано в ячейке (п. 4). Специальное устройство дает возможность с пульта управления машиной в любой момент работы выяснить содержимое любой ячейки, а также ввести дополнительные данные.

Каждая конструкция автоматических цифровых машин имеет свой набор команд и свои коды команд. В п. 6 мы приведем несколько простых примеров, на которых можно понять некоторые принципы составления программ, т. е. *программирования*; впрочем, в реальных машинах эти принципы реализуются экономнее. Более подробные указания можно найти, например, в книгах [7], [12], [14], [15], [24], [27], [31]. Составление серьезной программы является очень ответственным делом, часто занимает много времени и требует навыков работы.

6. Примеры программирования. Пусть требуется подсчитать решение системы уравнений первой степени

$$\left.\begin{aligned} ax+by&=m,\\ cx+dy&=n, \end{aligned}\right\}$$

причем будем считать значения чисел a, b, c, d, m, n заданными.

В силу формул (VI.2) искомое решение получается по формулам

$$x=\frac{md-bn}{ad-bc}, \quad y=\frac{an-mc}{ad-bc}. \tag{6}$$

Будем размещать команды в ячейках с номерами 1, 2, ...; мы пока не знаем, сколько потребуется команд. Шесть входных параметров a, b, c, d, m, n разместим соответственно в ячейках с номерами $\alpha+1$, $\alpha+2$, $\alpha+3$, $\alpha+4$, $\alpha+5$, $\alpha+6$; значение α мы уточним позже. Несколько дальнейших ячеек будут использованы в качестве *рабочих ячеек* для хранения промежуточных результатов; состояние этих ячеек к началу работы несущественно, так как при любой записи в ячейке предыдущая запись в ней автоматически стирается. Остальные ячейки ЗУ при составлении и работе программы не используются. Начнем с подсчета первого числителя (6). Чтобы вычислить md, требуется команда

1) $\times$ $\alpha+5$ $\alpha+4$ $\alpha+7$;

при этом мы перед командой для ориентировки пишем ее номер, который фактически нигде не пробивается. После исполнения этой команды в ячейке $(\alpha+7)$ появится число md. Следующая команда пусть имеет вид

2) $\times$ $\alpha+2$ $\alpha+6$ $\alpha+8$;

после ее исполнения в ячейке $(\alpha+8)$ появится число bn. Далее надо из числа md (ячейка $(\alpha+7)$) вычесть число bn (ячейка $(\alpha+8)$). Так как эти числа больше не понадобятся, то результат можно записать в ячейке $(\alpha+7)$ с помощью команды

3) $-$ $\alpha+7$ $\alpha+8$ $\alpha+7$;

конечно, здесь можно было использовать и ячейку $(\alpha+9)$, но в более сложных программах часто приходится экономить ячейки.

Аналогично находим знаменатель дроби (6):

4) $\times$ $\alpha+1$ $\alpha+4$ $\alpha+8$,
5) $\times$ $\alpha+2$ $\alpha+3$ $\alpha+9$,
6) $-$ $\alpha+8$ $\alpha+9$ $\alpha+8$.

Теперь надо числитель, хранящийся в ячейке $(\alpha+7)$, разделить на знаменатель, хранящийся в ячейке $(\alpha+8)$, и результат отпечатать:

7) : $\alpha+7$ $\alpha+8$ $\alpha+7$,
8) печ. $\alpha+7$;

по последней команде машина печатает содержимое ячейки, т. е. значение x, и переходит к исполнению следующей команды. Аналогично вычисляем y, имея в виду, что знаменатель для y уже вычис-

лен и хранится в ячейке $(\alpha+8)$

9)	×	$\alpha+1$	$\alpha+6$	$\alpha+7$,
10)	×	$\alpha+5$	$\alpha+3$	$\alpha+9$,
11)	—	$\alpha+7$	$\alpha+9$	$\alpha+7$,
12)	:	$\alpha+7$	$\alpha+8$	$\alpha+7$,
13)	печ.	$\alpha+7$,		
14)	ост.			

По 14-й команде машина останавливается. Итак, в нашей программе содержится 14 команд и потому можно принять $\alpha=14$. Тогда вся программа будет занимать 20 ячеек и иметь такой вид:

1)	×	19	18	21,
2)	×	16	20	22,
3)	—	21	22	21,
4)	×	15	18	22,
5)	×	16	17	23,
6)	—	22	23	22,
7)	:	21	22	21,
8)	печ.	21		,
9)	×	15	20	21,
10)	×	19	17	23,
11)	—	21	23	21,
12)	:	21	22	21,
13)	печ.	21,		
14)	ост.,			
15)	a,			
16)	b,			
17)	c,			
18)	d,			
19)	m,			
20)	n.			

При печатании на перфокартах можно было принять α бóльшим, например $\alpha=24$. Дело в том, что на каждой перфокарте пробивается по 12 строк (чисел или команд), и тогда бы все команды разместились на двух перфокартах, а все входные числа — на третьей, так как $\alpha+1=25$. Это дало бы возможность при изменившихся исходных данных заменить только третью карту, а первые две использовать повторно.

Итак, все промежуточные результаты в нашей программе требуют трех рабочих ячеек; при $\alpha=14$ — это ячейки № 21, 22 и 23. Теперь остается только пробить эту программу и пустить машину в ход. (На самом деле требуются еще некоторые команды, не имеющие для нас принципиального значения, например, команда ввода программы с перфокарт в ЗУ и команды перевода параметров из двоично-десятичной системы в двоичную.)

Теперь легко понять, что такое программа для вычисления значений функции (см. п. I.13). Пусть, например, рассматривается функция $y = x^2 - 3x + 7$. Проверьте, что тогда программа

1)	$\times$	α	α	8,
2)	$\times$	α	6	9,
3)	$-$	8	9	8,
4)	$+$	8	7	β,
5)	ост.,			
6)	3,			
7)	7			

ставит в соответствие любому значению x, помещенному в ячейку α, значение y, расположенное в ячейке β. Такая программа легко воспроизводима и, после пробивки, готова к работе. Имеются правила математических действий над такими программами: например, программы для интегрирования, нахождения экстремумов и т. п. Несомненно, что предстоит широкая замена формул на программы в многих разделах математики и ее приложений.

В разобранных примерах число команд в программе равнялось числу действий. Однако современные цифровые машины создавались в основном для вычислений, требующих сотен тысяч и миллионов действий. Ясно, что в этом случае писать отдельную команду на каждое действие невозможно. К счастью, при большой массе вычислений обычно многие промежуточные вычисления проводятся по одной и той же схеме несколько раз; а тогда в программе можно предусмотреть образование *циклов*, при которых УУ проходит один и тот же участок программы по нескольку раз. Циклы образуются с помощью команды *условной передачи управления*, которая имеет вид

$$\text{ПУ} \qquad N_1 \qquad N_2 \qquad N_3$$

(ПУ означает «передачу управления»; конечно, на самом деле надо указать код этой операции). Имеются различные варианты реализации этой команды. Мы будем для определенности считать, что по этой команде машина сравнивает содержимое $(N_1)'$ ячейки N_1 и содержимое $(N_2)'$ ячейки N_2; если $(N_1)' < (N_2)'$, то УУ переходит к исполнению очередной команды, если же $(N_1)' \geqslant (N_2)'$, то УУ переходит к исполнению команды, содержащейся в ячейке с номером N_3, причем в обоих случаях содержание ячеек не меняется. В частности, команда

$$\text{ПУ} \qquad 1 \qquad 1 \qquad N_3$$

означает, что УУ после прочтения ее обязательно переходит к исполнению команды, содержащейся в ячейке N_3; такую команду можно считать командой *безусловной передачи управления*.

Составим для примера программу для напечатания таблицы обратных величин к 1000 нат ральным числам, скажем, от 2001 до

3000; без команды передачи управления программа заняла бы очень большой объем, на деле же она получится очень маленькой. Поместим в ячейке $(\alpha+1)$ число 2000, в ячейке $(\alpha+2)$ число 1, а в ячейке $(\alpha+3)$ число 2999 (мы вскоре увидим, зачем оно нужно). Пусть первая команда имеет вид

1) $+ \quad \alpha+1 \quad \alpha+2 \quad \alpha+1.$

После выполнения ее в ячейке $(\alpha+1)$ взамен числа 2000 появится число 2001, для которого мы вычислим обратное и результат напечатаем:

2) $: \quad \alpha+2 \quad \alpha+1 \quad \alpha+4,$
3) печ. $\alpha+4.$

Сравним число, стоящее в ячейке $(\alpha+1)$, с 2999 и, поскольку оно еще не превзошло 2999, перейдем вновь к первой команде:

4) ПУ $\alpha+3 \quad \alpha+1 \quad 1.$

Так как в данном случае $(\alpha+3)'=2999$, $(\alpha+1)'=2001$, то, прочтя четвертую команду, УУ переходит к первой команде. После ее выполнения в ячейке $(\alpha+1)$ взамен 2001 появится 2002; по следующим двум командам будет подсчитано и напечатано обратное число к 2002. По четвертой команде машина сравнит 2999 с 2002 и вновь перейдет к исполнению первой команды и т. д. И только когда после очередного прибавления единицы в ячейке $(\alpha+1)$ появится число 3000 и обратная величина к 3000 будет отпечатана, четвертая команда «пропустит» УУ вперед, так как тогда будет $(\alpha+3)'=2999$, $(\alpha+1)'=3000$. Тогда остается остановить машину

5) ост.,

так как все результаты уже отпечатаны.

Итак, можно принять $\alpha=5$, и полная программа будет иметь вид

1)	+	6	7	6,
2)	:	7	6	9,
3)	печ.	9,		
4)	ПУ	8	6	1,
5)	ост.,			
6)	2000,			
7)	1,			
8)	2999.			

Для хранения промежуточных результатов нам понадобилась всего одна рабочая ячейка (№ 9); однако по мере работы содержимое шестой ячейки менялось от 2000 до 3000 через 1.

Приведем вариант, в котором к концу работы рассматриваемые обратные величины оказываются не отпечатанными, а размещенными в ячейках ЗУ с номерами от 10 до 1009, предполагая, что объем

ЗУ позволяет разместить столько величин (они могут быть использованы для каких-либо дальнейших вычислений):

1)	+	6	7	6,
2)	+	3	9	3,
3)	:	7	6	9,
4)	ПУ	8	6	1,
5)	ост.,			
6)	2000,			
7)	1,			
8)	2999,			
9)	(1 в последнем разряде ячейки, остальные 0).			

Здесь в девятую ячейку программы засылается *вспомогательное число*, не имеющее количественного значения, а служащее для *преобразования команд*, что является принципиально новым действием. Именно, после первого выполнения второй команды третья команда, воспринятая АУ как число, примет вид :7 6 10. (Заметим, что сложения по командам 1) и 2) выполняются по несколько различным правилам и потому должны иметь различные коды.) Поэтому после выполнения третьей команды 1:2001 отправится в 10-ю ячейку. После вторичного выполнения второй команды третья примет вид :7 6 11, и потому после вторичного выполнения третьей команды число 1:2002 отправится в 11-ю ячейку и т. д. Эта операция изменения адреса в команде называется *переадресацией*.

Таким образом, с командами можно совершать действия, автоматически преобразуя их по мере работы машины. Ясно, что это представляет дополнительные возможности в применении цифровых машин.

Команда условной передачи управления применяется и без циклов в тех случаях, когда требуется организовать *разветвление программы*, т. е. производить различные последовательности действий в зависимости от неизвестных заранее обстоятельств. Пусть, например, в процессе вычислений в определенной ячейке с номером β должно получиться число a и мы должны оставить его без изменения, если $a \geqslant 0$, или возвести в квадрат и результат оставить в ячейке β, если $a < 0$. Так как в ячейке с номером 0 обычно помещается число 0, то требуемое действие можно осуществить, поместив в соответствующем месте следующие команды:

. . . .

k)	ПУ	β	0	$k+2$,
$k+1$)	$\times$	β	β	β.

При этом разветвление осуществится автоматически, так что мы даже не будем знать, какой вариант был реализован, если не предусмотрим особо информацию об этом.

Покажем, наконец, пример программы, в которой общее число действий заранее не устанавливается. Пусть требуется решить ку-

бическое уравнение

$$x = 0{,}1x^3 + 1$$

методом итераций (п. V.3), начиная от значения $x_0 = 0$. Для этого разместим число 0 в ячейках $(\alpha + 1)$ и $(\alpha + 2)$ (соответствующие строки перфокарты просто не пробиваются), число 0,1 — в ячейке $(\alpha + 3)$ и число 1 — в ячейке $(\alpha + 4)$. Последовательные приближения будем помещать в ячейке $(\alpha + 2)$.

Вычисление последующего приближения по предыдущему осуществляется по командам:

1)	$\times$	$\alpha + 2$	$\alpha + 2$	$\alpha + 5$,
2)	$\times$	$\alpha + 5$	$\alpha + 2$	$\alpha + 5$,
3)	$\times$	$\alpha + 3$	$\alpha + 5$	$\alpha + 5$,
4)	$+$	$\alpha + 5$	$\alpha + 4$	$\alpha + 5$.

Итак, последующее приближение помещено в ячейку $(\alpha + 5)$. Его надо сравнить с предыдущим, содержащимся в ячейке $(\alpha + 2)$, — и если они не совпадают, поместить последующее приближение в ячейку $(\alpha + 2)$ и повторить итерацию; если же приближения совпадают, то надо отпечатать результат. Это можно сделать, например, по командам (проверьте!):

5)	$\lvert - \rvert$	$\alpha + 2$	$\alpha + 5$	$\alpha + 2$

(по этой команде абсолютная величина разности содержимого ячеек $(\alpha + 2)$ и $(\alpha + 5)$ засылается в ячейку $(\alpha + 2)$),

6)	ПУ	$\alpha + 1$	$\alpha + 2$	9,
7)	$+$	$\alpha + 5$	$\alpha + 1$	$\alpha + 2$,
8)	ПУ	1	1	1,
9)	печ.	$\alpha + 5$,		
10)	ост.			

Итак, можно положить $\alpha = 10$ и выписать всю программу, занимающую 14 ячеек. По этой программе машина будет совершать итерации до тех пор, пока последующее приближение не совпадет с предыдущим (это возможно потому, что результаты вычислений автоматически округляются), после чего она напечатает результат и остановится. Отметим, что если процесс итераций не сходится, то произойдет либо переполнение ячеек, либо *зацикливание*, при котором машина сама не может остановиться и для ее останова требуется вмешательство извне с пульта управления.

Программы для более сложных задач могут занимать большой объем; однако они нередко включают в себя более простые, часто повторяющиеся задачи, например вычисление значения синуса от участвующих величин и т. п. Решение этих более простых задач осуществляется с помощью *стандартных подпрограмм*, т. е. набора команд, которые составляются заранее и хранятся в определенных

ячейках оперативного ЗУ или во внешнем ЗУ; тогда при составлении программы для более сложной задачи обращение к стандартной подпрограмме производится по одной команде.

Для облегчения программирования в последние годы разработано несколько *универсальных машинных языков*, на которых программа расписывается в более привычных математических терминах, чем это было описано выше. Программа на таком языке не зависит от типа применяемой машины и печатается на устройстве, аналогичном пишущей машинке; затем специальный *трансля́тор* автоматически переводит эту программу в программу той машины, которая будет применяться. У нас наиболее известен разработанный в 1960 г. язык *Алгол-60* (от английского algorithmical language — алгоритмический язык), освоение которого сделает применение вычислительных машин значительно более доступным. Впрочем, опытный программист, учитывая специфику конкретной машины, обычно может составить программу более экономно без применения универсального языка, что бывает существенно при необходимости экономить *машинное время*, т. е. время работы машины.

Ошибки в вычислениях могут получиться из-за неправильного составления или пробивки программы либо же по вине самой машины; последние могут быть систематическими (например, из-за выхода из строя отдельных элементов) или случайными (*сбои*, при которых элемент сам собой переходит из одного состояния в другое, например, из-за внезапных помех). Систематические ошибки выявляются с помощью регулярного решения контрольных задач с известными ответами. Правильность составления программы проверяется при ее *отладке*, которая производится перед работой машины. При отладке через машину пропускаются отдельные участки программы; тогда же производится грубая прикидка промежуточных результатов, важная для того, чтобы эти результаты умещались в ячейках ЗУ (п. 4). Для контроля пробивки программы применяется повторная пробивка с автоматической сверкой пробитых карт друг с другом. Для выявления сбоев применяются некоторые средства контроля правильности вычислений, известные из арифметики, а в более ответственных расчетах применяют повторное вычисление; кроме того, вероятность сбоя стараются уменьшить при помощи тех или иных конструктивных особенностей машины при ее проектировании, доводке и модификациях.

7. Краткие сведения о советских электронных универсальных цифровых машинах. Приведем в заключение некоторые сведения о серийных советских электронных универсальных цифровых машинах. Следует иметь в виду, что эти данные являются сугубо ориентировочными, а порой и устарелыми, так как по некоторым конструкциям выпускаются модификации, имеющие более высокие характеристики. Кроме того, имеется ряд марок специализированных машин (п. 2), а также малосерийные машины, обладающие интерес-

ными конструктивными особенностями и хорошими показателями (*Арагац*, *Киев*, *Сéтунь* и другие).

Первая советская электронная универсальная цифровая вычислительная машина (*ЭЦВМ*), получившая наименование *М*-3, была создана в 1952 г. под руководством советского ученого, специалиста по электроэнергетическим системам, электрическим и математическим машинам И. С. Брука. Она работала по двухадресной системе с фиксированной запятой и имела скромные по нынешним оценкам характеристики: средняя скорость действия 30 операций в секунду (в дальнейшем была доведена до 1500), ввод и вывод данных со скоростью 1 числа в 2 секунды, ЗУ на магнитном барабане емкостью в 1024 ячейки по 31 разряду. Вскоре этот же коллектив создал машину *М*-2, модифицированную в 1954—1955 гг. и обладающую более высокими характеристиками: быстродействие 2000 операций/*сек*, трехадресная система с плавающей или фиксированной запятой, внутреннее ЗУ на магнитных сердечниках — 4096 ячеек по 34 разряда, внешнее ЗУ на 200 000 чисел на магнитной ленте.

В 1953 г. в АН СССР под руководством советского ученого-электротехника С. А. Лебедева была создана более крупная машина *БЭСМ* со средним быстродействием в 7000 операций/*сек*. (Все приводимые данные по быстродействию являются лишь ориентировочными, так как время выполнения различных действий существенно различное.) Вскоре появились машины *Урал* (100 операций/*сек*), *Стрела* (2000 операций/*сек*, трехадресная система с плавающей запятой, внутреннее ЗУ на электронно-лучевых трубках — 2048 ячеек по 43 разряда, внешнее ЗУ на 100 000 чисел) и другие.

«*Первое поколение*» советских ЭЦВМ было, как и во всем мире, основано на применении электронных вакуумных ламп: их количество для указанных машин находилось в диапазоне от 774 (*М*-3) до 5851 (*Стрела*). Сравнительно большие размеры ламп не только приводят к тому, что машина потребляет большую мощность и размещается на большой площади (29 *квт* и 22 $м^2$ для *М*-2, 124 *квт* и 400 $м^2$ для *Стрелы* и т. п.), но и из-за возникновения паразитных емкостей существенно ограничивают быстродействие. Поэтому в 60-е годы постепенно перешли на производство ЭЦВМ *второго поколения*, основанных на полупроводниковых диодах и триодах (транзисторах), хотя в эксплуатации имеется еще большое число ламповых машин. Машины второго поколения занимают гораздо меньше места, потребляют меньшую мощность и гораздо более устойчивы в работе, чем соответствующие машины первого поколения, в частности не требуют круглосуточного включения. Впрочем, качество ЭЦВМ зависит не только от свойств ее элементов, но и от многих ее конструктивных и схемных особенностей. Отметим, что в последние годы производится опытная разработка ЭЦВМ *третьего поколения*, основанных на последних достижениях электроники — твердых схемах и схемах на тонких пленках.

Так как серийные машины все время совершенствуются, то сейчас имеется несколько «семейств» ЭЦВМ, имеющих общность в наименовании. Широко известно семейство *БЭСМ*, о первом представителе которого мы уже упоминали. Машины этого семейства работают по трехадресной системе с плавающей запятой, причем среди ламповых машин они наиболее устойчивы в работе (имеют наименьший процент сбоев). *БЭСМ*-2*М* имеет быстродействие 8000—10 000 операций/*сек*, внутреннее ЗУ на магнитных сердечниках — 2048 ячеек по 39 разрядов, промежуточное ЗУ на двух магнитных барабанах по 5720 чисел со скоростью выборки 800 чисел/*сек*, внешнее ЗУ на магнитной ленте—120 000 чисел со скоростью выборки 400 чисел/*сек*. Ввод с помощью перфоленты со скоростью 15 чисел/*сек* или с перфокарт, вывод на печатающее устройство с той же скоростью. Имеет 5000 электронных ламп, потребляет мощность 75 *квт*, размещается на площади 150—200 *м*². Следующая модификация, *БЭСМ*-3*М* (и близкая к ней *БЭСМ*-4), принадлежит уже второму поколению ЭЦВМ. Она имеет среднее быстродействие 20 000 операций/*сек*, внутреннее ЗУ от 4096 до 8192 ячеек по 45 разрядов, промежуточное ЗУ на четырех барабанах по 65 536 ячеек со скоростью выборки 12 000 чисел/*сек*, внешнее ЗУ—4 000 000 чисел со скоростью выборки 5000 чисел/*сек*. Имеет 15 000 транзисторов и 25 000 полупроводниковых диодов, потребляет мощность 7 *квт*, занимает площадь 50 *м*². В 1967 г. пущена новая модификация, *БЭСМ*-6, со средним быстродействием 1 000 000 операций/*сек*. Эта машина одновременно решает несколько задач, причем специальная *программа-диспетчер* во время действия по какой-либо задаче устройств ввода и вывода, работающих медленнее, чем АУ, автоматически переключает АУ на решение другой задачи. Особенностью *БЭСМ*-6 является также наличие *буферной памяти*, расположенной между ЗУ и АУ, из которой информация извлекается в 10 раз быстрее, чем из основного ЗУ.

Другое семейство ЭЦВМ начала машина *М*-20. Она работает по трехадресной системе с плавающей запятой и со средним быстродействием 20 000 операций в секунду. Оперативное ЗУ на ферритовых сердечниках содержит 4096 ячеек по 45 разрядов с временем выборки ячейки 6 *мксек*. Внешняя память—три магнитных барабана общей емкостью в 12 288 ячеек, время выборки или записи 70 *мксек*, и четыре десятидорожные магнитные ленты емкостью свыше 300 000 ячеек, с автоматическим поиском требуемой зоны и средним временем выборки или записи 300 *мксек*. Ввод данных с перфокарт со скоростью 20 строк/*сек*, результаты печатаются со скоростью 15—20 строк/*сек* или перфорируются со скоростью 10 строк/*сек*; результаты могут также выводиться на магнитный барабан одновременно с вычислениями. Машина имеет 4500 электронных ламп и около 35 000 полупроводниковых диодов, потребляет мощность 50 *квт* без охлаждения и вентиляции и занимает площадь около 370 *м*². Ее модификация *М*-220 принадлежит второму поколению ЭЦВМ, имеет среднее быстродействие 25 000 операций/*сек*, внутреннее ЗУ от 4096 до 16 384 ячеек, общую емкость барабанов 65 536 ячеек и лент от 4 до 16 миллионов ячеек, скорость ввода с перфокарт до 140 строк/*сек*, потребляет мощность 7—10 *квт*, занимает площадь 60—70 *м*².

ЭЦВМ *Урал*, упомянутая выше, также была родоначальницей семейства, для которого характерно применение одноадресной системы. В 1959 г. машина была модернизирована и получила наименование *Урал*-2. При этом оперативное ЗУ было выполнено на ферритовых сердечниках, что позволило значительно повысить быстродействие (в среднем до 5000 операций в секунду). Объем оперативного ЗУ повысился до 2048 ячеек по 40 разрядов. Внешнее ЗУ выполнено на магнитном барабане объемом в 8196 ячеек со скоростью считывания 3000 чисел в секунду и на магнитной ленте объемом в 100 000 ячеек со скоростью считывания 1000 чисел в секунду (не считая времени на перемещение требуемого участка ленты под считывающую головку). Ввод данных с перфоленты со скоростью 150 чисел в секунду. Машина имеет 2100 электронных ламп, потребляет мощность 25 *квт* и раз-

мещается на площади 150—200 *м*². Модификации *Урал*-4 и *Урал*-11 принадлежат второму поколению ЭЦВМ. Последняя имеет быстродействие 30 000 операций/*сек*, причем к машине можно подключить одновременно до 16 устройств ввода и вывода; результаты вычислений могут фиксироваться на перфокартах, перфо- и магнитных лентах или непосредственно поступать к регуляторам производственных процессов, передаваться по каналам связи, печататься на бумажной ленте или регистрироваться самописцами.

Семейства *Минск* и *Раздан* с самого начала принадлежали второму поколению ЭЦВМ. *Минск*-2 имеет быстродействие 5—6 тысяч операций в секунду, работает по двухадресной системе с плавающей или фиксированной запятой. Оперативное ЗУ на магнитных сердечниках имеет 4096 (8192 для *Минск*-22) ячеек по 37 разрядов, время считывания 24 *мксек*. Внешняя память на магнитной ленте от 400 до 1600 тысяч ячеек, время выборки или записи 400 *мксек*. Ввод с перфоленты со скоростью 800 строк в секунду, вывод на печать 20 строк в секунду или на перфорацию 20 (80 для *Минск*-22) строк в секунду. Занимает площадь 40—50 *м*², потребляет мощность 4 *квт*. В 1968 г. создана машина *Минск*-32 с быстродействием 30 000 операций/*сек*.

Раздан-2 имеет основной агрегат, по размерам ненамного превышающий кухонный буфет Машина предназначена для исследовательских лабораторий, конструкторских бюро, организаций, производящих обработку статистической и экономической информации. Работает со скоростью 5000 операций в секунду. Оперативное ЗУ имеет 2048 ячеек по 36 разрядов, внешняя память на магнитных барабанах общей емкостью 120 000 ячеек. Ввод со скоростью 35 строк в секунду, печать со скоростью 20 строк в секунду. Питается от сети переменного тока, потребляет 3 *квт*.

Совсем малые размеры — несколько больше письменного стола — имеют машины *Мир*, *Проминь*, *Наири*, предназначенные для решения инженерных, научных, экономических задач на основе универсального машинного языка (п. 7). Отметим также машины *Днепр* и *УМ*-1, предназначенные для контроля и управления производственными процессами, расчета оптимальных режимов; здесь сигналы в машину могут поступать от датчиков различного типа, а управляющие сигналы, выработанные машиной, передаются на регуляторы.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Арлей Н. и Бух К. Р., Введение в теорию вероятностей и математическую статистику, ИЛ, 1951.

2. Бараненков Г. С. и др., Задачи и упражнения по математическому анализу для втузов под ред. Б. П Демидовича, изд. 5, «Наука», 1966.

3. Березин И. С. и Жидков Н. П., Методы вычислений, т. 1, «Наука», 1966; т. 2, Физматгиз, 1959.

4. Берман Г. Н., Сборник задач по курсу математического анализа, изд. 14, «Наука», 1965

5. Бермант А. Ф. и Араманович И. Г., Краткий курс магематического анализа для втузов, изд. 4, «Наука», 1966.

6. Бронштейн И. Н. и Семендяев К. А., Справочник по математике для инженеров и учащихся втузов, изд. 10, «Наука», 1965.

7. Брудно А. Л., Введение в программирование (в содержательных обозначениях), «Наука», 1965.

8. Бут Э., Численные методы, Физматгиз, 1959.

9. Вентцель Е. С., Теория вероятностей, изд. 3, «Наука», 1964.

10. Гельфанд И. М., Лекции по линейной алгебре, изд. 3, «Наука», 1966.

11. Гнеденко Б. В., Курс теории вероятностей, изд. 4, «Наука», 1965.

12. Гнеденко Б. В., Королюк В. С. и Ющенко Е. Л., Элементы программирования, Физматгиз, 1963.

13. Градштейн И. С. и Рыжик И. М., Таблицы интегралов, сумм, рядов и произведений, изд. 4, «Наука», 1963.

14. Гутер Р. С., Арлазаров В. Л. и Усков А. В., Практика программирования (справочник), «Наука», 1965.

15. Гутер Р. С., Овчинский Б. В., Резниковский П. Т., Программирование и вычислительная математика, «Наука», 1965.

16. Демидович Б П., Сборник задач и упражнений по математическому анализу, изд. 6, «Наука», 1966.

17. Демидович Б. П. и Марон И. А., Основы вычислительной математики, изд. 3, «Наука», 1966.

18. Демидович Б. П., Марон И. А., Шувалова Э. З., Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения, изд. 2, «Наука», 1963.

19. Ефимов Н. В., Краткий курс аналитической геометрии, изд. 8, «Наука», 1965.

20. Загускин В. Л., Справочник по численным методам решения алгебраических и трансцендентных уравнений, Физматгиз, 1960.

21. Зельдович Я. Б., Мышкис А. Д., Элементы прикладной математики, «Наука», 1965.

22. Камке Э., Справочник по обыкновенным дифференциальным уравнениям, перев. с нем. под ред. С. В. Фомина, «Наука», 1965.

23. Камке Э., Справочник по дифференциальным уравнениям в частных производных первого порядка, «Наука», 1966.

24. Китов А. И. и Криницкий Н. А., Электронные цифровые машины и программирование, изд. 2, Физматгиз, 1961.
25 Клетеник Д. В., Сборник задач по аналитической геометрии, изд. 8, «Наука», 1965.
26. Коллатц Л., Численные методы решения дифференциальных уравнений, ИЛ, 1953.
27. Криницкий Н. А., Миронов Г. А. и Фролов Г. Д., Программирование, изд. 2, «Наука», 1966.
28. Крылов А. Н., Лекции о приближенных вычислениях, Гостехиздат, 1954.
29. Лебедев Н. Н., Специальные функции и их приложения, изд. 2, «Наука», 1963.
30 Левин Л., Методы решений технических задач на аналоговых вычислительных машинах, «Мир», 1966.
31. Мак-Кракен, Программирование для цифровых вычислительных машин, ИЛ, 1960.
32. Маркс К. и Энгельс Ф., Сочинения, изд. 2, т. 20.
33. «Математический практикум» под редакцией Г. Н. Положего, Физматгиз, 1960.
34. Милн В., Численное решение дифференциальных уравнений, ИЛ, 1955.
35. Милн В., Численный анализ, ИЛ, 1951.
36. Мишина А. П. и Проскуряков И. В., Высшая алгебра (серия «Справочная математическая библиотека»), изд. 2, «Наука», 1965.
37. Панов Д. Ю., Счетная линейка, изд. 18, «Наука», 1966.
38. Пискунов Н. С., Дифференциальное и интегральное исчисления для втузов, тт. 1, 2, «Наука», 1966.
39. Романовский П. И., Ряды Фурье. Теория поля. Аналитические и специальные функции. Преобразование Лапласа, изд 4, «Наука», 1964.
40. Румшиский Л. З., Элементы теории вероятностей, изд. 2, «Наука», 1963.
41. Сальвадори М., Численные методы в технике, ИЛ, 1955.
42. Семендяев К. А., Счетная линейка. Краткое руководство, изд. 8, Физматгиз, 1957.
43. Смирнов В. И., Курс высшей математики, тт. 1, 2, «Наука», 1965; т. 3, ч. 1, изд. 7, Физматгиз, 1956.
44. Смирнов Н. В., Дунин-Барковский И. В., Курс теории вероятностей и математической статистики для технических приложений, Физматгиз, 1959.
45. Смолянский М. Л., Таблицы неопределенных интегралов, изд. 4, «Наука», 1967.
46. Толстов Г. П., Элементы математического анализа, т. 1, «Наука», 1965; т. 2, «Наука», 1966.
47. Фаддеев Д. К. и Фаддеева В. Н., Вычислительные методы линейной алгебры, изд. 2, «Наука», 1963.
48 Фихтенгольц Г. М., Курс дифференциального и интегрального исчисления, тт. I—III, «Наука», 1966.
49 Фукс Б. А. и Шабат Б. В., Функции комплексного переменного и некоторые их приложения, изд. 3, «Наука», 1964.
50. Хаусхолдер А., Основы численного анализа, ИЛ, 1956.
51. Цубербиллер О. Н., Задачи и упражнения по аналитической геометрии, изд. 28, «Наука», 1966.
52. Этерман И. И., Математические машины непрерывного действия, Машгиз, 1957.
53. Янке Е. и Эмде Ф., Таблицы функций с формулами и кривыми, Физматгиз, 1959.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

УКАЗАТЕЛЬ ОБОЗНАЧЕНИЙ